

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>7</sup> :</b> C12N 15/12, C07K 14/705, 16/28, C12Q 1/68, G01N 33/53, 33/68, A61K 38/16, A01K 67/027	<b>A1</b>	<b>(11) International Publication Number:</b> WO 00/12707 <b>(43) International Publication Date:</b> 9 March 2000 (09.03.00)
<b>(21) International Application Number:</b> PCT/US99/20084 <b>(22) International Filing Date:</b> 2 September 1999 (02.09.99) <b>(30) Priority Data:</b> 09/145,745 2 September 1998 (02.09.98) US 09/383,745 26 August 1999 (26.08.99) US <b>(71) Applicant:</b> MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US). <b>(72) Inventors:</b> GLUCKSMANN, Maria, Alexandra; 33 Summit Road, Lexington, MA 02173 (US). SILOS-SANTIAGO, Inmaculada; 18 Hilliard Street, Cambridge, MA 02138 (US). <b>(74) Agents:</b> BROWN, Anne et al.; Alston & Bird LLP, P.O. Drawer 34009, Charlotte, NC 28234-4009 (US).		<b>(81) Designated States:</b> AE, AL, AM, AT, AT (Utility model), AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, CZ (Utility model), DE, DE (Utility model), DK, DK (Utility model), DM, EE, EE (Utility model), ES, FI, FI (Utility model), GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SK (Utility model), SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
<b>(54) Title:</b> 14926 RECEPTOR, A NOVEL G-PROTEIN COUPLED RECEPTOR		
<b>(57) Abstract</b> <p>The present invention relates to a newly identified receptor belonging to the superfamily of G-protein-coupled receptors. The invention also relates to polynucleotides encoding the receptor. The invention further relates to methods using the receptor polypeptides and polynucleotides as a target for diagnosis and treatment in receptor-mediated disorders. The invention further relates to drug-screening methods using the receptor polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the receptor polypeptides and polynucleotides. The invention further relates to procedures for producing the receptor polypeptides and polynucleotides.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## 14926 RECEPTOR, A NOVEL G-PROTEIN COUPLED RECEPTOR

## FIELD OF THE INVENTION

The present invention relates to a newly identified receptor belonging to the superfamily of G-protein-coupled receptors. The invention also relates to polynucleotides encoding the receptor. The invention further relates to methods using the receptor polypeptides and polynucleotides as a target for diagnosis and treatment in receptor-mediated disorders. The invention further relates to drug-screening methods using the receptor polypeptides and polynucleotides to identify agonists and antagonists for diagnosis and treatment. The invention further encompasses agonists and antagonists based on the receptor polypeptides and polynucleotides. The invention further relates to procedures for producing the receptor polypeptides and polynucleotides.

## BACKGROUND OF THE INVENTION

G-protein coupled receptors

G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for transducing a signal within a cell. GPCRs have three structural domains: an amino terminal extracellular domain, a transmembrane domain containing seven transmembrane segments, three extracellular loops, and three intracellular loops, and a carboxy terminal intracellular domain. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins and effectors (intracellular enzymes and channels modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular inputs.

GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al.*, *J. Clin. Invest.* 92:1119-1125 (1993); McKusick *et al.*, *J. Med. Genet.* 30:1-26 (1993)). Specific defects in the rhodopsin gene and the V2

vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosum (Nathans *et al.*, *Annu. Rev. Genet.* 26:403-424(1992)), and nephrogenic diabetes insipidus (Holtzman *et al.*, *Hum. Mol. Genet.* 2:1201-1204 (1993)). These receptors are of critical importance to both the central nervous system and peripheral physiological processes. Evolutionary analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the  $\beta$ 2-adrenergic receptor and currently represented by over 200 unique members (Dohlman *et al.*, *Annu. Rev. Biochem.* 60:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.*, *Science* 254:1024-1026 (1991); Lin *et al.*, *Science* 254:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, *Science* 258 597:603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al.*, *Science* 241:1467-1472 (1988)); and Family V, the fungal mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem.* 61:1097-1129 (1992)).

There are also a small number of other proteins which present seven putative hydrophobic segments and appear to be unrelated to GPCRs; they have not been shown to couple to G-proteins. *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein which has been extensively studied and does not show evidence of being a GPCR (Hart *et al.*, *Proc. Natl. Acad. Sci. USA* 90:5047-5051 (1993)). The gene *frizzled* (*fz*) in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, *fz* has not been shown to couple to G-proteins (Vinson *et al.*, *Nature* 338:263-264 (1989)).

G proteins represent a family of heterotrimeric proteins composed of  $\alpha$ ,  $\beta$  and  $\gamma$  subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors containing seven transmembrane segments. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the  $\alpha$ -subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the  $\beta\gamma$ -subunits. The GTP-bound form of the  $\alpha$ -subunit typically functions as an effector-modulating moiety, leading to the

production of second messengers, such as cAMP (e.g., by activation of adenylyl cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of  $\alpha$ -subunits are known in humans. These subunits associate with a smaller pool of  $\beta$  and  $\gamma$  subunits. Examples of mammalian G proteins include Gi, Go, Gq, Gs and Gt. G proteins are described extensively in Lodish *et al.*, *Molecular Cell Biology*, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference. GPCRs, G proteins and G protein-linked effector and second messenger systems have been reviewed in *The G-Protein Linked Receptor Fact Book*, Watson *et al.*, eds., Academic Press (1994).

GPCRs are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs. The present invention advances the state of the art by providing a previously unidentified human GPCR.

## SUMMARY OF THE INVENTION

It is an object of the invention to identify novel GPCRs.

It is a further object of the invention to provide novel GPCR polypeptides that are useful as reagents or targets in receptor assays applicable to treatment and diagnosis of GPCR-mediated disorders.

It is a further object of the invention to provide polynucleotides corresponding to the novel GPCR receptor polypeptides that are useful as targets and reagents in receptor assays applicable to treatment and diagnosis of GPCR-mediated disorders and useful for producing novel receptor polypeptides by recombinant methods.

A specific object of the invention is to identify compounds that act as agonists and antagonists and modulate the expression of the novel receptor.

A further specific object of the invention is to provide compounds that modulate expression of the receptor for treatment and diagnosis of GPCR-related disorders.

The invention is thus based on the identification of a novel GPCR, designated the 14926 receptor.

The invention provides isolated 14926 receptor polypeptides including a polypeptide having the amino acid sequence shown in SEQ ID NO 1.

The invention also provides isolated 14926 receptor nucleic acid molecules having the sequence shown in SEQ ID NO 2.

5     The invention also provides variant polypeptides having an amino acid sequence that is substantially homologous to the amino acid sequence shown in SEQ ID NO 1.

The invention also provides variant nucleic acid sequences that are substantially homologous to the nucleotide sequence shown in SEQ ID NO 2.

10     The invention also provides fragments of the polypeptide shown in SEQ ID NO 1 and nucleotide shown in SEQ ID NO 2, as well as substantially homologous fragments of the polypeptide or nucleic acid.

The invention also provides vectors and host cells for expressing the receptor nucleic acid molecules and polypeptides and particularly recombinant  
15     vectors and host cells.

The invention also provides methods of making the vectors and host cells and methods for using them to produce the receptor nucleic acid molecules and polypeptides.

The invention also provides antibodies that selectively bind the receptor  
20     polypeptides and fragments.

The invention also provides methods of screening for compounds that modulate the activity of the receptor polypeptides. Modulation can be at the level of the polypeptide receptor or at the level of controlling the expression of nucleic acid (RNA or DNA) expressing the receptor polypeptide.

25     The invention also provides a process for modulating receptor polypeptide activity, especially using the screened compounds, including to treat conditions related to expression of the receptor polypeptides.

The invention also provides diagnostic assays for determining the presence of and level of the receptor polypeptides or nucleic acid molecules in a biological  
30     sample.

The invention also provides diagnostic assays for determining the presence of a mutation in the receptor polypeptides or nucleic acid molecules.

## DESCRIPTION OF THE DRAWINGS

5        **Figure 1** shows the 14926 nucleotide sequence (SEQ ID NO 2) and the deduced 14926 amino acid sequence (SEQ ID NO 1). It is predicted that amino acids 1-23 constitute the amino terminal extracellular domain, amino acids 24-341 constitute the region spanning the transmembrane domain, and amino acids 342-370 constitute the carboxy terminal intracellular domain. The transmembrane domain contains seven transmembrane segments, three extracellular loops and three intracellular loops. The transmembrane segments are found from about amino acid 24 to about amino acid 46, from about amino acid 56 to about amino acid 78, from about amino acid 96 to about amino acid 117, from about amino acid 133 to about amino acid 154, from about amino acid 185 to about amino acid 209, 10 from about amino acid 286 to about amino acid 307, and from about amino acid 318 to about amino acid 341. Within the region spanning the entire transmembrane domain are three intracellular and three extracellular loops. The three intracellular loops are found from about amino acid 47 to about amino acid 55, from about amino acid 118 to about amino acid 132, and from about amino acid 210 to about amino acid 285. The three extracellular loops are found at from about amino acid 79 to about amino acid 95, from about amino acid 155 to about amino acid 184, and from about amino acid 308 to about amino acid 317.

15        The transmembrane domain includes a GPCR signal transduction signature, TRY, at residues 118-120. The sequence includes an arginine at residue 119, an invariant amino acid in GPCRs.

20        **Figure 2** shows a comparison of the 14926 receptor against the Prosite data base of protein patterns, specifically showing a high score against the seven transmembrane segment rhodopsin superfamily. The underlined area shows a GPCR signature, and specifically the position of an arginine residue, conserved in 30 GPCRs. The most commonly conserved sequence is an aspartate, arginine, tyrosine (DRY) triplet. DRY is implicated in signal transduction. Arginine is invariant. Aspartate is conservatively placed in several GPCRs. In the present

case, the arginine is found in the sequence TRY, which matches the position of DRY or invariant arginine in GPCRs of the rhodopsin superfamily of receptors.

Figure 3 shows an analysis of the 14926 amino acid sequence:  $\alpha$ turn and coil regions; hydrophilicity; amphipathic regions; flexible regions; antigenic index; and surface probability plot.

Figure 4 shows a 14926 receptor hydrophobicity plot. The amino acids correspond to 1-370 and show the seven transmembrane segments.

10

Figure 5 shows an analysis of the 14926 open reading frame for amino acids corresponding to specific functional sites. A glycosylation site is found at amino acids 3-6, which corresponds to the amino terminal extracellular domain. A second glycosylation site is found at amino acids 83-86, which corresponds to the first extracellular loop. A third glycosylation site is found at amino acids 182-185, which spans the second extracellular loop and fifth transmembrane segment. A fourth glycosylation site is found at amino acids 227-230, which corresponds to the third intracellular loop. A fifth glycosylation site occurs at amino acids 264-267, also in the third intracellular loop. A cyclic AMP or cyclic GMP-dependent protein kinase phosphorylation site is found at amino acids 131-134 and spans the second intracellular loop and fourth transmembrane segment, and at amino acids 281-284, corresponding to the third intracellular loop. A protein kinase C phosphorylation site is found at amino acids 80-82, corresponding to the first intracellular loop. A second protein kinase C phosphorylation site is found at amino acids 93-95, corresponding to the first extracellular loop. A third protein kinase C phosphorylation site is found at amino acids 130-132, corresponding to the second intracellular loop. A fourth protein kinase C phosphorylation site is found at amino acids 178-180, corresponding to the second extracellular loop. A fifth protein kinase C phosphorylation site is found at amino acids 266-268, corresponding to the third intracellular loop. A sixth protein kinase C phosphorylation site is found at amino acids 342-344, corresponding to the carboxy terminal intracellular domain. A casein kinase II phosphorylation site occurs at amino acids 342-345, corresponding to the carboxy terminal intracellular

25

30



domain. N-myristoylation sites occur at amino acids 84-89 and 90-95, corresponding to the first extracellular loop; 101-106, corresponding to the third transmembrane segment; 237-242 and 258-263, corresponding to the third intracellular loop; and 318-323, corresponding to the seventh transmembrane segment. An amidation site is found at amino acids 266-269, corresponding to the third intracellular loop. In addition, amino acids corresponding in position to the GPCR signature and containing the invariant arginine are found in the sequence TRY at amino acids 118-120.

## 10 DETAILED DESCRIPTION OF THE INVENTION

### Receptor function/signal pathway

The 14926 receptor protein is a GPCR that participates in signaling pathways. As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR (14926 protein). Examples of such functions include mobilization of intracellular molecules that participate in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP<sub>2</sub>), inositol 1,4,5-triphosphate (IP<sub>3</sub>) and adenylate cyclase; polarization of the plasma membrane; production or secretion of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival. Since the 14926 receptor protein is expressed in brain cells participating in a 14926 receptor protein signaling pathway include, but are not limited to cells derived from these tissues.

The response mediated by the receptor protein depends on the type of cell. For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the receptor protein, it is universal that the protein is a GPCR and interacts with G proteins to produce one or more secondary signals, in a variety of intracellular

signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell.

As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP<sub>2</sub>) as well as to the activities of these molecules. PIP<sub>2</sub> is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP<sub>2</sub> to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP<sub>3</sub>). Once formed IP<sub>3</sub> can diffuse to the endoplasmic reticulum surface where it can bind an IP<sub>3</sub> receptor, e.g., a calcium channel protein containing an IP<sub>3</sub> binding site. IP<sub>3</sub> binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP<sub>3</sub> can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP<sub>4</sub>), a molecule which can cause calcium entry into the cytoplasm from the extracellular medium. IP<sub>3</sub> and IP<sub>4</sub> can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-bisphosphate (IP<sub>2</sub>) and inositol 1,3,4-triphosphate, respectively. These inactive products can be recycled by the cell to synthesize PIP<sub>2</sub>. The other second messenger produced by the hydrolysis of PIP<sub>2</sub>, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF-kB. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP<sub>2</sub> or one of its metabolites.

Another signaling pathway in which the receptor may participate is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand-induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR

can lead to the activation of the enzyme adenylyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.

## 10 Polypeptides

The invention is based on the identification of a novel G-coupled protein receptor. Specifically, an expressed sequence tag (EST) was selected based on homology to G-protein-coupled receptor sequences. This EST was used to design primers based on sequences that it contains and used to identify a cDNA from a human brain cDNA library. Positive clones were sequenced and the overlapping fragments were assembled. Analysis of the assembled sequence revealed that the cloned cDNA molecule encodes a G-protein coupled receptor.

The invention thus relates to a novel GPCR having the deduced amino acid sequence shown in Figure 1 (SEQ ID NO 1).

20 The "14926 receptor polypeptide" or "14926 receptor protein" refers to the polypeptide in SEQ ID NO 1. The term "receptor protein" or "receptor polypeptide", however, further includes the numerous variants described herein, as well as fragments derived from the full length 14926 polypeptide and variants.

The present invention thus provides an isolated or purified 14926 receptor polypeptide and variants and fragments thereof.

25 The 14926 polypeptide is a 370 residue protein exhibiting three main structural domains. The amino terminal extracellular domain is identified to be within residues 1 to about 23 in SEQ ID NO 1. The transmembrane domain is identified to be within residues from about 24 to about 341 in SEQ ID NO 1. The carboxy terminal intracellular domain is identified to be within residues from about 342 to 370 in SEQ ID NO 1. The transmembrane domain contains seven segments that span the membrane. The transmembrane segments are found from about amino acid 24 to about amino acid 46, from about amino acid 56 to about amino

acid 78, from about amino acid 96 to about amino acid 117, from about amino acid 133 to about amino acid 154, from about amino acid 185 to about amino acid 209, from about amino acid 286 to about amino acid 307, and from about amino acid 318 to about amino acid 341. Within the region spanning the entire  
5 transmembrane domain are three intracellular and three extracellular loops. The three intracellular loops are found from about amino acid 47 to about amino acid 55, from about amino acid 118 to about amino acid 132, and from about amino acid 210 to about amino acid 285. The three extracellular loops are found at from about amino acid 79 to about amino acid 95, from about amino acid 155 to about  
10 amino acid 184, and from about amino acid 308 to about amino acid 317.

A glycosylation site is found at amino acids 3-6, which corresponds to the amino terminal extracellular domain. A second glycosylation site is found at amino acids 83-86, which corresponds to the first extracellular loop. A third glycosylation site is found at amino acids 182-185, which spans the second  
15 extracellular loop and fifth transmembrane segment. A fourth glycosylation site is found at amino acids 227-230, which corresponds to the third intracellular loop. A fifth glycosylation site occurs at amino acids 264-267, also in the third intracellular loop. A cyclic AMP or cyclic GMP-dependent protein kinase phosphorylation site is found at amino acids 131-134 and spans the second intracellular loop and fourth  
20 transmembrane segment, and at amino acids 281-284, corresponding to the third intracellular loop. A protein kinase C phosphorylation site is found at amino acids 80-82, corresponding to the first intracellular loop. A second protein kinase C phosphorylation site is found at amino acids 93-95, corresponding to the first extracellular loop. A third protein kinase C phosphorylation site is found at amino  
25 acids 130-132, corresponding to the second intracellular loop. A fourth protein kinase C phosphorylation site is found at amino acids 178-180, corresponding to the second extracellular loop. A fifth protein kinase C phosphorylation site is found at amino acids 266-268, corresponding to the third intracellular loop. A sixth protein kinase C phosphorylation site is found at amino acids 342-344,  
30 corresponding to the carboxy terminal intracellular domain. A casein kinase II phosphorylation site occurs at amino acids 342-345, corresponding to the carboxy terminal intracellular domain. N-myristoylation sites occur at amino acids 84-89 and 90-95, corresponding to the first extracellular loop; 101-106, corresponding to

the third transmembrane segment; 237-242 and 258-263, corresponding to the third intracellular loop; and 318-323, corresponding to the seventh transmembrane segment. An amidation site is found at amino acids 266-269, corresponding to the third intracellular loop.

5

The transmembrane domain includes a GPCR signal transduction signature, TRY, at residues 118-120. The sequence includes an arginine at residue 119, an invariant amino acid in GPCRs.

Based on a GenBank search, homology was shown to serotonin receptors.

10

As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. A polypeptide, however, can be joined to another polypeptide with which it is not normally associated in a cell and still be considered "isolated" or "purified."

15

The receptor polypeptides can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful and considered to contain an isolated form of the polypeptide. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity.

20

In one embodiment, the language "substantially free of cellular material" includes preparations of the receptor polypeptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the receptor polypeptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the protein preparation.

25

The language "substantially free of chemical precursors or other chemicals" includes preparations of the receptor polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis.

30

In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals. In one embodiment, the receptor polypeptide comprises the amino acid sequence shown in SEQ ID NO 1. However, the invention also encompasses sequence variants. Variants include a substantially homologous protein encoded by the same genetic locus in an organism, i.e., an allelic variant. The receptor maps to chromosome 7, in close proximity to marker Bda06f04. Variants also encompass proteins derived from other genetic loci in an organism, but having substantial homology to the 14926 receptor protein of SEQ ID NO 1. Variants also include proteins substantially homologous to the 14926 receptor protein but derived from another organism, i.e., an ortholog. Variants also include proteins that are substantially homologous to the 14926 receptor protein that are produced by chemical synthesis. Variants also include proteins that are substantially homologous to the 14926 receptor protein that are produced by recombinant methods. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

As used herein, two proteins (or a region of the proteins) are substantially homologous when the amino acid sequences are at least about 55-60%, 60-65%, 65-70%, typically at least about 70-75%, more typically at least about 80-85%, and most typically at least about 90-95% or more homologous. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence hybridizing to the nucleic acid sequence, or portion thereof, of the sequence shown in SEQ ID NO 2 under stringent conditions as more fully described below.

To determine the percent homology of two amino acid sequences, or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one protein or nucleic acid for optimal alignment with the other protein or nucleic acid). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in one sequence is occupied by the same amino acid

residue or nucleotide as the corresponding position in the other sequence, then the molecules are homologous at that position. As used herein, amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity". The percent homology between the two sequences is a function of the number of identical positions shared by the sequences (i.e., per cent homology equals the number of identical positions/total number of positions times 100).

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by the 14926 polypeptide. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

TABLE 1: Conservative Amino Acid Substitutions.

Aromatic	Phenylalanine
	Tryptophan
	Tyrosine
Hydrophobic	Leucine
	Isoleucine
	Valine
Polar	Glutamine
	Asparagine
Basic	Arginine
	Lysine
	Histidine
Acidic	Aspartic Acid
	Glutamic Acid
Small	Alanine
	Serine
	Threonine
	Methionine
	Glycine

Both identity and similarity can be readily calculated (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; 5 *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and 10 *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCG program package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F. *et al.*, *J. Molec. Biol.* 215:403 (1990)).



A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these.

Variant polypeptides can be fully functional or can lack function in one or more activities. Thus, in the present case, variations can affect the function, for example, of one or more of the regions corresponding to ligand binding, membrane association, G-protein binding and signal transduction.

Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids which result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

As indicated, variants can be naturally-occurring or can be made by recombinant means or chemical synthesis to provide useful and novel characteristics for the receptor polypeptide. This includes preventing immunogenicity from pharmaceutical formulations by preventing protein aggregation.

Useful variations further include alteration of ligand binding characteristics. For example, one embodiment involves a variation at the binding site that results in binding but not release, or slower release, of ligand. A further useful variation at the same sites can result in a higher affinity for ligand. Useful variations also include changes that provide for affinity for another ligand. Another useful variation includes one that allows binding but which prevents activation by the ligand. Another useful variation includes variation in the transmembrane G-protein-binding/signal transduction domain that provides for reduced or increased binding by the appropriate G-protein or for binding by a different G-protein than the one with which the receptor is normally associated. Another useful variation provides a fusion protein in which one or more domains or subregions is operationally fused to one or more domains or subregions from another G-protein coupled receptor.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule.

5 The resulting mutant molecules are then tested for biological activity such as receptor binding or *in vitro*, or *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al.* *Science* 255:306-312 (1992)).

10 Substantial homology can be to the entire nucleic acid or amino acid sequence or to fragments of these sequences.

The invention thus also includes polypeptide fragments of the 14926 receptor protein. Fragments can be derived from the amino acid sequence shown in SEQ ID NO 1. However, the invention also encompasses fragments of the  
15 variants of the 14926 receptor protein as described herein.

The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed prior to the present invention.

Fragments can retain one or more of the biological activities of the protein,  
20 for example the ability to bind to a G-protein or ligand. Fragments can also be useful as an immunogen to generate receptor antibodies.

Biologically active fragments can comprise a domain or motif, e.g., an extracellular or intracellular domain or loop, one or more transmembrane segments, or parts thereof, G-protein binding site, or GPCR signature,  
25 glycosylation sites, cAMP and cGMP-dependent, protein kinase C, and casein kinase II phosphorylation sites, amidation and myristoylation sites. Such peptides can be, for example, 7, 10, 15, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length.

Possible fragments include, but are not limited to: 1) soluble peptides  
30 comprising the entire amino terminal extracellular domain from amino acid 1 to about amino acid 23 of SEQ ID NO 1, or parts thereof; 2) peptides comprising the entire carboxy terminal intracellular domain from about amino acid 342 to amino acid 370 of SEQ ID NO 1, or parts thereof; 3) peptides comprising the region

spanning the entire transmembrane domain from about amino acid 24 to about amino acid 341, or parts thereof; 4) any of the specific transmembrane segments, or parts thereof, from about amino acid 24 to about amino acid 46, from about amino acid 56 to about amino acid 78, from about amino acid 96 to about amino acid 117, from about amino acid 133 to about amino acid 154, from about amino acid 185 to about amino acid 209, from about amino acid 286 to about amino acid 307, and from about amino acid 318 to about amino acid 341; 5) any of the three intracellular or three extracellular loops, or parts thereof, from about amino acid 79 to about amino acid 95, from about amino acid 155 to about amino acid 184, from about amino acid 308 to about amino acid 317, from about amino acid 47 to about amino acid 55, from about amino acid 118 to about amino acid 132, and from about amino acid 210 to about amino acid 285. Fragments further include combinations of the above fragments, such as an amino terminal domain combined with one or more transmembrane segments and the attendant extra or intracellular loops or one or more transmembrane segments, and the attendant intra or extracellular loops, plus the carboxy terminal domain. Thus, any of the above fragments can be combined. Other fragments include the mature protein from about amino acid 6 to 370. Other fragments contain the various functional sites described herein, such as phosphorylation sites, glycosylation sites, and myristoylation sites and a sequence containing the GPCR signature sequence. Fragments, for example, can extend in one or both directions from the functional site to encompass 5, 10, 15, 20, 30, 40, 50, or up to 100 amino acids. Further, fragments can include sub-fragments of the specific domains mentioned above, which sub-fragments retain the function of the domain from which they are derived. Fragments also include amino acid sequences greater than 135 amino acids. Fragments also include antigenic fragments and specifically those shown to have a high antigenic index in Figure 3. Further specific fragments include a fragment from about 136 to about 169 and subfragments thereof greater than about 5 amino acids; a fragment including any of the sequences from about 1-120 but extending beyond about 120; a fragment including any of the sequences from about 1-135 but extending beyond about 135; from about 304-370 and subfragments thereof greater than about 7 amino acids; from about 320-370 and subfragments thereof greater than about 7 amino acids; from about 326-370 and

subfragments thereof greater than about 7 amino acids; a fragment from around 350-370 and subfragments thereof greater than about 5 amino acids; a fragment containing sequences from around 170-304 but including sequences either extending beyond about 170 or 304 or both.

5 Accordingly, possible fragments include fragments defining a ligand-binding site, fragments defining a glycosylation site, fragments defining membrane association, fragments defining phosphorylation sites, fragments defining sites of interaction with G proteins and signal transduction, and fragments defining myristoylation sites. By this is intended a discrete fragment that provides the  
10 relevant function or allows the relevant function to be identified. In a preferred embodiment, the fragment contains the ligand-binding site.

The invention also provides fragments with immunogenic properties. These contain an epitope-bearing portion of the 14926 receptor protein and variants. These epitope-bearing peptides are useful to raise antibodies that bind  
15 specifically to a receptor polypeptide or region or fragment. These peptides can contain at least 7, 12, 14, or between at least about 15 to about 30 amino acids. Regions having a high antigenicity index are shown in Figure 3.

Non-limiting examples of antigenic polypeptides that can be used to generate antibodies include peptides derived from the amino terminal extracellular  
20 domain or any of the intra or extracellular loops.

The epitope-bearing receptor and polypeptides may be produced by any conventional means (Houghten, R.A., *Proc. Natl. Acad. Sci. USA* 82:5131-5135 (1985)). Simultaneous multiple peptide synthesis is described in U.S. Patent No. 4,631,211.

25 Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the receptor fragment and an additional region fused to  
30 the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion proteins. These comprise a receptor protein operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the receptor protein. "Operatively

linked" indicates that the receptor protein and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the receptor protein.

5 In one embodiment the fusion protein does not affect receptor function per se. For example, the fusion protein can be a GST-fusion protein in which the receptor sequences are fused to the C-terminus of the GST sequences. Other types of fusion proteins include, but are not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions and Ig fusions. Such fusion proteins, particularly poly-His fusions, can  
10 facilitate the purification of recombinant receptor protein. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence. Therefore, in another embodiment, the fusion protein contains a heterologous signal sequence at its N-terminus.

15 EP-A-O 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP-A 0232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify  
20 antagonists. Bennett *et al.* (*J. Mol. Recog.* 8:52-58 (1995)) and Johanson *et al.* (*J. Biol. Chem.* 270, 16:9459-9471 (1995)). Thus, this invention also encompasses soluble fusion proteins containing a receptor polypeptide and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE). Preferred as immunoglobulin is the constant part of  
25 the heavy chain of human IgG, particularly IgG1, where fusion takes place at the hinge region. For some uses it is desirable to remove the Fc after the fusion protein has been used for its intended purpose, for example when the fusion protein is to be used as antigen for immunizations. In a particular embodiment, the Fc part can be removed in a simple way by a cleavage sequence which is also incorporated  
30 and can be cleaved with factor Xa.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional

techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A receptor protein-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the receptor protein.

Another form of fusion protein is one that directly affects receptor functions. Accordingly, a receptor polypeptide is encompassed by the present invention in which one or more of the receptor domains (or parts thereof) has been replaced by homologous domains (or parts thereof) from another G-protein coupled receptor or other type of receptor. Accordingly, various permutations are possible. The amino terminal extracellular domain, or subregion thereof, (for example, ligand-binding) can be replaced with the domain or subregion from another ligand-binding receptor protein. Alternatively, the entire transmembrane domain, or any of the seven segments or loops, or parts thereof, for example, G-protein-binding/signal transduction, can be replaced. Finally, the carboxy terminal intracellular domain or subregion can be replaced. Thus, chimeric receptors can be formed in which one or more of the native domains or subregions has been replaced.

The isolated receptor protein can be purified from cells that naturally express it, such as from brain, spleen, lung, kidney, skeletal muscle, fetal liver, adult liver, heart, and K562 (erythroblast/erythroleukemia), 293 (kidney), and Jurkat (T cell) cell lines, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods.

In one embodiment, the protein is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the receptor polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard

protein purification techniques. Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally-occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in polypeptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art. Accordingly, the polypeptides also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence for purification of the mature polypeptide or a pro-protein sequence.

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by

Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

As is also well known, polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of post-translation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translational natural processes and by synthetic methods.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. Blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally-occurring and synthetic polypeptides. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

The modifications can be a function of how the protein is made. For recombinant polypeptides, for example, the modifications will be determined by the host cell posttranslational modification capacity and the modification signals in the polypeptide amino acid sequence. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cells often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to efficiently express mammalian proteins having native patterns of glycosylation. Similar considerations apply to other modifications.

The same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain more than one type of modification.

### Polypeptide uses

The receptor polypeptides are useful for producing antibodies specific for the 14926 receptor protein, regions, or fragments. Regions having a high antigenicity index score are shown in Figure 3.



The receptor polypeptides (including those variants and fragments which may have been disclosed prior to the present invention) are useful for biological assays related to GPCRs. Such assays involve any of the known GPCR functions or activities or properties useful for diagnosis and treatment of GPCR-related conditions.

The receptor polypeptides are useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the receptor protein, as a biopsy or expanded in cell culture. In one embodiment, however, cell-based assays involve recombinant host cells expressing the receptor protein.

The receptor polypeptides can be used to identify compounds that modulate receptor activity. Both 14926 protein and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the receptor. These compounds can be further screened against a functional receptor to determine the effect of the compound on the receptor activity. Compounds can be identified that activate (agonist) or inactivate (antagonist) the receptor to a desired degree.

The receptor polypeptides can be used to screen a compound for the ability to stimulate or inhibit interaction between the receptor protein and a target molecule that normally interacts with the receptor protein. The target can be ligand or a component of the signal pathway with which the receptor protein normally interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol turnover and/or adenylate cyclase, or phospholipase C activation). The assay includes the steps of combining the receptor protein with a candidate compound under conditions that allow the receptor protein or fragment to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the receptor protein and the target, such as any of the associated effects of signal transduction such as G-protein phosphorylation, cyclic AMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide

libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g.,  
5 Songyang *et al.*, *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')<sub>2</sub>, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

10 One candidate compound is a soluble full-length receptor or fragment that competes for ligand binding. Other candidate compounds include mutant receptors or appropriate fragments containing mutations that affect receptor function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow  
15 release, is encompassed by the invention.

The invention provides other end points to identify compounds that modulate (stimulate or inhibit) receptor activity. The assays typically involve an assay of events in the signal transduction pathway that indicate receptor activity. Thus, the expression of genes that are up- or down-regulated in response to the  
20 receptor protein dependent signal cascade can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase. Alternatively, phosphorylation of the receptor protein, or a receptor protein target, could also be measured.

Binding and/or activating compounds can also be screened by using  
25 chimeric receptor proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a G-protein-binding region can  
30 be used that interacts with a different G-protein than that which is recognized by the native receptor. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. Alternatively, the entire transmembrane portion or subregions (such as transmembrane segments or

intracellular or extracellular loops) can be replaced with the entire transmembrane portion or subregions specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding region are derived. This allows for assays to be performed in other than the specific host cell from which the receptor is derived. Alternatively, the amino terminal extracellular domain (and/or other ligand-binding regions) could be replaced by a domain (and/or other binding region) binding a different ligand, thus, providing an assay for test compounds that interact with the heterologous amino terminal extracellular domain (or region) but still cause signal transduction. Finally, activation can be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

The receptor polypeptides are also useful in competition binding assays in methods designed to discover compounds that interact with the receptor. Thus, a compound is exposed to a receptor polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble receptor polypeptide is also added to the mixture. If the test compound interacts with the soluble receptor polypeptide, it decreases the amount of complex formed or activity from the receptor target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the receptor. Thus, the soluble polypeptide that competes with the target receptor region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is desirable to immobilize either the receptor protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase/14926 fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., <sup>35</sup>S-labeled) and the candidate compound, and the mixture incubated under conditions

conductive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated  
5 from the matrix, separated by SDS-PAGE, and the level of receptor-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the  
10 protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a receptor-binding protein and a candidate compound are incubated in the receptor protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such  
15 complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the receptor protein target molecule, or which are reactive with receptor protein and compete with the target molecule; as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

20 Modulators of receptor protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the receptor pathway, by treating cells that express the 14926 protein, such as in brain, spleen, lung, skeletal muscle, kidney, liver, and heart.

Disorders involving the spleen include, but are not limited to, splenomegaly,  
25 including nonspecific acute splenitis, congestive splenomegaly, and splenic infarcts; neoplasms, congenital anomalies, and rupture. Disorders associated with splenomegaly include infections, such as nonspecific splenitis, infectious mononucleosis, tuberculosis, typhoid fever, brucellosis, cytomegalovirus, syphilis, malaria, histoplasmosis, toxoplasmosis, kala-azar, trypanosomiasis, schistosomiasis,  
30 leishmaniasis, and echinococcosis; congestive states related to partial hypertension, such as cirrhosis of the liver, portal or splenic vein thrombosis, and cardiac failure; lymphohematogenous disorders, such as Hodgkin disease, non-Hodgkin lymphomas/leukemia, multiple myeloma, myeloproliferative disorders, hemolytic

anemias, and thrombocytopenic purpura; immunologic-inflammatory conditions, such as rheumatoid arthritis and systemic lupus erythematosus; storage diseases such as Gaucher disease, Niemann-Pick disease, and mucopolysaccharidoses; and other conditions, such as amyloidosis, primary neoplasms and cysts, and secondary neoplasms.

Disorders involving the lung include, but are not limited to, congenital anomalies; atelectasis; diseases of vascular origin, such as pulmonary congestion and edema, including hemodynamic pulmonary edema and edema caused by microvascular injury, adult respiratory distress syndrome (diffuse alveolar damage), pulmonary embolism, hemorrhage, and infarction, and pulmonary hypertension and vascular sclerosis; chronic obstructive pulmonary disease, such as emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis; diffuse interstitial (infiltrative, restrictive) diseases, such as pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia (pulmonary infiltration with eosinophilia), *Bronchiolitis obliterans*-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, including Goodpasture syndrome, idiopathic pulmonary hemosiderosis and other hemorrhagic syndromes, pulmonary involvement in collagen vascular disorders, and pulmonary alveolar proteinosis; complications of therapies, such as drug-induced lung disease, radiation-induced lung disease, and lung transplantation; tumors, such as bronchogenic carcinoma, including paraneoplastic syndromes, bronchioloalveolar carcinoma, neuroendocrine tumors, such as bronchial carcinoid, miscellaneous tumors, and metastatic tumors; pathologies of the pleura, including inflammatory pleural effusions, noninflammatory pleural effusions, pneumothorax, and pleural tumors, including solitary fibrous tumors (pleural fibroma) and malignant mesothelioma.

Disorders involving the liver include, but are not limited to, hepatic injury; jaundice and cholestasis, such as bilirubin and bile formation; hepatic failure and cirrhosis, such as cirrhosis, portal hypertension, including ascites, portosystemic shunts, and splenomegaly; infectious disorders, such as viral hepatitis, including hepatitis A-E infection and infection by other hepatitis viruses, clinicopathologic syndromes, such as the carrier state, asymptomatic infection, acute viral hepatitis, chronic viral hepatitis, and fulminant hepatitis; autoimmune hepatitis; drug- and

toxin-induced liver disease, such as alcoholic liver disease; inborn errors of metabolism and pediatric liver disease, such as hemochromatosis, Wilson disease,  $\alpha_1$ -antitrypsin deficiency, and neonatal hepatitis; intrahepatic biliary tract disease, such as secondary biliary cirrhosis, primary biliary cirrhosis, primary sclerosing cholangitis, and anomalies of the biliary tree; circulatory disorders, such as impaired blood flow into the liver, including hepatic artery compromise and portal vein obstruction and thrombosis, impaired blood flow through the liver, including passive congestion and centrilobular necrosis and peliosis hepatis, hepatic vein outflow obstruction, including hepatic vein thrombosis (Budd-Chiari syndrome) and veno-occlusive disease; hepatic disease associated with pregnancy, such as preeclampsia and eclampsia, acute fatty liver of pregnancy, and intrahepatic cholestasis of pregnancy; hepatic complications of organ or bone marrow transplantation, such as drug toxicity after bone marrow transplantation, graft-versus-host disease and liver rejection, and nonimmunologic damage to liver allografts; tumors and tumorous conditions, such as nodular hyperplasias, adenomas, and malignant tumors, including primary carcinoma of the liver and metastatic tumors.

Disorders involving the brain include, but are limited to, disorders involving neurons, and disorders involving glia, such as astrocytes, oligodendrocytes, ependymal cells, and microglia; cerebral edema, raised intracranial pressure and herniation, and hydrocephalus; malformations and developmental diseases, such as neural tube defects, forebrain anomalies, posterior fossa anomalies, and syringomyelia and hydromyelia; perinatal brain injury; cerebrovascular diseases, such as those related to hypoxia, ischemia, and infarction, including hypotension, hypoperfusion, and low-flow states--global cerebral ischemia and focal cerebral ischemia--infarction from obstruction of local blood supply, intracranial hemorrhage, including intracerebral (intraparenchymal) hemorrhage, subarachnoid hemorrhage and ruptured berry aneurysms, and vascular malformations, hypertensive cerebrovascular disease, including lacunar infarcts, slit hemorrhages, and hypertensive encephalopathy; infections, such as acute meningitis, including acute pyogenic (bacterial) meningitis and acute aseptic (viral) meningitis, acute focal suppurative infections, including brain abscess, subdural empyema, and extradural abscess, chronic bacterial meningoencephalitis, including tuberculosis and mycobacterioses, neurosyphilis, and neuroborreliosis (Lyme disease), viral

meningoencephalitis, including arthropod-borne (Arbo) viral encephalitis, *Herpes simplex* virus Type 1, *Herpes simplex* virus Type 2, *Varicella-zoster* virus (*Herpes zoster*), cytomegalovirus, poliomyelitis, rabies, and human immunodeficiency virus 1, including HIV-1 meningoencephalitis (subacute encephalitis), vacuolar myelopathy, AIDS-associated myopathy, peripheral neuropathy, and AIDS in children, progressive multifocal leukoencephalopathy, subacute sclerosing panencephalitis, fungal meningoencephalitis, other infectious diseases of the nervous system; transmissible spongiform encephalopathies (prion diseases); demyelinating diseases, including multiple sclerosis, multiple sclerosis variants, acute disseminated encephalomyelitis and acute necrotizing hemorrhagic encephalomyelitis, and other diseases with demyelination; degenerative diseases, such as degenerative diseases affecting the cerebral cortex, including Alzheimer disease and Pick disease, degenerative diseases of basal ganglia and brain stem, including Parkinsonism, idiopathic Parkinson disease (paralysis agitans), progressive supranuclear palsy, corticobasal degeneration, multiple system atrophy, including striatonigral degeneration, Shy-Drager syndrome, and olivopontocerebellar atrophy, and Huntington disease; spinocerebellar degenerations, including spinocerebellar ataxias, including Friedreich ataxia, and ataxia-telangiectasia, degenerative diseases affecting motor neurons, including amyotrophic lateral sclerosis (motor neuron disease), bulbospinal atrophy (Kennedy syndrome), and spinal muscular atrophy; inborn errors of metabolism, such as leukodystrophies, including Krabbe disease, metachromatic leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, and Canavan disease, mitochondrial encephalomyopathies, including Leigh disease and other mitochondrial encephalomyopathies; toxic and acquired metabolic diseases, including vitamin deficiencies such as thiamine (vitamin B<sub>1</sub>) deficiency and vitamin B<sub>12</sub> deficiency, neurologic sequelae of metabolic disturbances, including hypoglycemia, hyperglycemia, and hepatic encephatopathy, toxic disorders, including carbon monoxide, methanol, ethanol, and radiation, including combined methotrexate and radiation-induced injury; tumors, such as gliomas, including astrocytoma, including fibrillary (diffuse) astrocytoma and glioblastoma multiforme, pilocytic astrocytoma, pleomorphic xanthoastrocytoma, and brain stem glioma, oligodendroglioma, and ependymoma and related paraventricular mass lesions, neuronal tumors, poorly differentiated neoplasms, including medulloblastoma, other parenchymal tumors,

including primary brain lymphoma, germ cell tumors, and pineal parenchymal tumors, meningiomas, metastatic tumors, paraneoplastic syndromes, peripheral nerve sheath tumors, including schwannoma, neurofibroma, and malignant peripheral nerve sheath tumor (malignant schwannoma), and neurocutaneous syndromes (phakomatoses), including neurofibromatosis, including Type 1 neurofibromatosis (NF1) and TYPE 2 neurofibromatosis (NF2), tuberous sclerosis, and Von Hippel-Lindau disease.

Disorders involving the heart, include but are not limited to, heart failure, including but not limited to, cardiac hypertrophy, left-sided heart failure, and right-sided heart failure; ischemic heart disease, including but not limited to angina pectoris, myocardial infarction, chronic ischemic heart disease, and sudden cardiac death; hypertensive heart disease, including but not limited to, systemic (left-sided) hypertensive heart disease and pulmonary (right-sided) hypertensive heart disease; valvular heart disease, including but not limited to, valvular degeneration caused by calcification, such as calcific aortic stenosis, calcification of a congenitally bicuspid aortic valve, and mitral annular calcification, and myxomatous degeneration of the mitral valve (mitral valve prolapse), rheumatic fever and rheumatic heart disease, infective endocarditis, and noninfected vegetations, such as nonbacterial thrombotic endocarditis and endocarditis of systemic lupus erythematosus (Libman-Sacks disease), carcinoid heart disease, and complications of artificial valves; myocardial disease, including but not limited to dilated cardiomyopathy, hypertrophic cardiomyopathy, restrictive cardiomyopathy, and myocarditis; pericardial disease, including but not limited to, pericardial effusion and hemopericardium and pericarditis, including acute pericarditis and healed pericarditis, and rheumatoid heart disease; neoplastic heart disease, including but not limited to, primary cardiac tumors, such as myxoma, lipoma, papillary fibroelastoma, rhabdomyoma, and sarcoma, and cardiac effects of noncardiac neoplasms; congenital heart disease, including but not limited to, left-to-right shunts--late cyanosis, such as atrial septal defect, ventricular septal defect, patent ductus arteriosus, and atrioventricular septal defect, right-to-left shunts--early cyanosis, such as tetralogy of fallot, transposition of great arteries, truncus arteriosus, tricuspid atresia, and total anomalous pulmonary venous connection, obstructive congenital anomalies, such as coarctation of aorta, pulmonary stenosis.



and atresia, and aortic stenosis and atresia, and disorders involving cardiac transplantation.

Disorders involving the kidney include, but are not limited to, congenital anomalies including, but not limited to, cystic diseases of the kidney, that include but are not limited to, cystic renal dysplasia, autosomal dominant (adult) polycystic kidney disease, autosomal recessive (childhood) polycystic kidney disease, and cystic diseases of renal medulla, which include, but are not limited to, medullary sponge kidney, and nephronophthisis-uremic medullary cystic disease complex, acquired (dialysis-associated) cystic disease, such as simple cysts; glomerular diseases including pathologies of glomerular injury that include, but are not limited to, in situ immune complex deposition, that includes, but is not limited to, anti-GBM nephritis, Heymann nephritis, and antibodies against planted antigens, circulating immune complex nephritis, antibodies to glomerular cells, cell-mediated immunity in glomerulonephritis, activation of alternative complement pathway, epithelial cell injury, and pathologies involving mediators of glomerular injury including cellular and soluble mediators, acute glomerulonephritis, such as acute proliferative (poststreptococcal, postinfectious) glomerulonephritis, including but not limited to, poststreptococcal glomerulonephritis and nonstreptococcal acute glomerulonephritis, rapidly progressive (crescentic) glomerulonephritis, nephrotic syndrome, membranous glomerulonephritis (membranous nephropathy), minimal change disease (lipoid nephrosis), focal segmental glomerulosclerosis, membranoproliferative glomerulonephritis, IgA nephropathy (Berger disease), focal proliferative and necrotizing glomerulonephritis (focal glomerulonephritis), hereditary nephritis, including but not limited to, Alport syndrome and thin membrane disease (benign familial hematuria), chronic glomerulonephritis, glomerular lesions associated with systemic disease, including but not limited to, systemic lupus erythematosus, Henoch-Schönlein purpura, bacterial endocarditis, diabetic glomerulosclerosis, amyloidosis, fibrillary and immunotactoid glomerulonephritis, and other systemic disorders; diseases affecting tubules and interstitium, including acute tubular necrosis and tubulointerstitial nephritis, including but not limited to, pyelonephritis and urinary tract infection, acute pyelonephritis, chronic pyelonephritis and reflux nephropathy, and tubulointerstitial nephritis induced by drugs and toxins, including but not limited to, acute drug-induced

interstitial nephritis, analgesic abuse nephropathy, nephropathy associated with nonsteroidal anti-inflammatory drugs, and other tubulointerstitial diseases including, but not limited to, urate nephropathy, hypercalcemia and nephrocalcinosis, and multiple myeloma; diseases of blood vessels including benign nephrosclerosis, malignant hypertension and accelerated nephrosclerosis, renal artery stenosis, and thrombotic microangiopathies including, but not limited to, classic (childhood) hemolytic-uremic syndrome, adult hemolytic-uremic syndrome/thrombotic thrombocytopenic purpura, idiopathic HUS/TTP, and other vascular disorders including, but not limited to, atherosclerotic ischemic renal disease, atheroembolic renal disease, sickle cell disease nephropathy, diffuse cortical necrosis, and renal infarcts; urinary tract obstruction (obstructive uropathy); urolithiasis (renal calculi, stones); and tumors of the kidney including, but not limited to, benign tumors, such as renal papillary adenoma, renal fibroma or hamartoma (renomedullary interstitial cell tumor), angiomyolipoma, and oncocytoma, and malignant tumors, including renal cell carcinoma (hypernephroma, adenocarcinoma of kidney), which includes urothelial carcinomas of renal pelvis.

Disorders involving the skeletal muscle include tumors such as rhabdomyosarcoma.

These methods of treatment include the steps of administering the modulators of protein activity in a pharmaceutical composition as described herein, to a subject in need of such treatment.

Expression in specific cell lines suggest that the receptor is involved in cellular growth and proliferation as well as in inflammation. Accordingly, disorders that are specifically relevant include those related to dysfunctional growth and proliferation, such as hyperplasia, especially tumor growth, and to the treatment of inflammatory conditions. Preferred disorders include those involving the brain.

The receptor polypeptides also are useful to provide a target for diagnosing a disease or predisposition to disease mediated by the receptor protein, especially in the disorders and tissues discussed above, relevant to treatment. Accordingly, methods are provided for detecting the presence, or levels of, the receptor protein in a cell, tissue, or organism. The method involves contacting a biological sample

with a compound capable of interacting with the receptor protein such that the interaction can be detected.

One agent for detecting receptor protein is an antibody capable of selectively binding to receptor protein. A biological sample includes tissues, cells  
5 and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The receptor protein also provides a target for diagnosing active disease, or predisposition to disease, in a patient having a variant receptor protein. Thus, receptor protein can be isolated from a biological sample, assayed for the presence  
10 of a genetic mutation that results in aberrant receptor protein. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered receptor activity in cell-based or cell-free assay, alteration in ligand or  
15 antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein.

*In vitro* techniques for detection of receptor protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and  
20 immunofluorescence. Alternatively, the protein can be detected *in vivo* in a subject by introducing into the subject a labeled anti-receptor antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods which detect the allelic variant of a receptor protein expressed in a subject  
25 and methods which detect fragments of a receptor protein in a sample.

The receptor polypeptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M., *Clin. Exp. Pharmacol. Physiol.* 23(10-11)  
30 :983-985 (1996), and Linder, M.W., *Clin. Chem.* 43(2):254-266 (1997). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual

can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the receptor protein in which one or more of the receptor functions in one population is different from those in another population. The polypeptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic polypeptides could be identified.

The receptor polypeptides are also useful for monitoring therapeutic effects during clinical trials and other treatment. Thus, the therapeutic effectiveness of an agent that is designed to increase or decrease gene expression, protein levels or receptor activity can be monitored over the course of treatment using the receptor polypeptides as an end-point target.

The receptor polypeptides are also useful for treating a receptor-associated disorder. Accordingly, methods for treatment include the use of soluble receptor or fragments of the receptor protein that compete for ligand binding. These receptors or fragments can have a higher affinity for the ligand so as to provide effective competition.

### Antibodies

The invention also provides antibodies that selectively bind to the 14926 receptor protein and its variants and fragments. An antibody is considered to selectively bind, even if it also binds to other proteins that are not substantially  
5 homologous with the receptor protein. These other proteins share homology with a fragment or domain of the receptor protein. This conservation in specific regions gives rise to antibodies that bind to both proteins by virtue of the homologous sequence. In this case, it would be understood that antibody binding to the receptor protein is still selective.

10 To generate antibodies, an isolated receptor polypeptide is used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. Either the full-length protein or antigenic peptide fragment can be used. Regions having a high antigenicity index are shown in Figure 3. Antibodies are preferably prepared from these regions or from  
15 discrete fragments in these regions. However, antibodies can be prepared from any region of the peptide as described herein. A preferred fragment produces an antibody that diminishes or completely prevents ligand-binding. Antibodies can be developed against the entire receptor or portions of the receptor, for example, the intracellular carboxy terminal domain, the amino terminal extracellular domain, the  
20 entire transmembrane domain or specific segments, any of the intra or extracellular loops, or any portions of the above. Antibodies may also be developed against specific functional sites, such as the site of ligand-binding, the site of G protein coupling, or sites that are phosphorylated, glycosylated, or myristoylated.

An antigenic fragment will typically comprise at least 7-12 contiguous  
25 amino acid residues. The antigenic peptide can comprise, however, at least 14 amino acid residues, at least 15 amino acid residues, at least 20 amino acid residues, or at least 30 amino acid residues. In one embodiment, fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions. These fragments are not to be construed, however, as  
30 encompassing any fragments which may be disclosed prior to the invention.

Antibodies can be polyclonal or monoclonal. An intact antibody, or a fragment thereof (e.g. Fab or F(ab')<sub>2</sub>) can be used.

Detection can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase,  $\beta$ -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include  $^{125}\text{I}$ ,  $^{131}\text{I}$ ,  $^{35}\text{S}$  or  $^3\text{H}$ .

An appropriate immunogenic preparation can be derived from native, recombinantly expressed, protein or chemically synthesized peptides.

#### Antibody Uses

The antibodies can be used to isolate a receptor protein by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural receptor protein from cells and recombinantly produced receptor protein expressed in host cells.

The antibodies are useful to detect the presence of receptor protein in cells or tissues to determine the pattern of expression of the receptor among various tissues in an organism and over the course of normal development.

The antibodies can be used to detect receptor protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. The antibodies can be used to assess abnormal tissue distribution or abnormal expression during development.

Antibody detection of circulating fragments of the full length receptor protein can be used to identify receptor turnover.

Further, the antibodies can be used to assess receptor expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to receptor function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, or level

of expression of the receptor protein, the antibody can be prepared against the normal receptor protein. If a disorder is characterized by a specific mutation in the receptor protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant receptor protein.

5       The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Antibodies can be developed against the whole receptor or portions of the receptor, for example, portions of the amino terminal extracellular domain or extracellular loops.

10       The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting receptor expression level or the presence of aberrant receptors and aberrant tissue distribution or developmental expression, antibodies directed against the receptor or relevant fragments can be used to monitor therapeutic efficacy.

15       Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic receptor proteins can be used to identify individuals that require modified treatment modalities.

20       The antibodies are also useful as diagnostic tools as an immunological marker for aberrant receptor protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

      The antibodies are also useful for tissue typing. Thus, where a specific receptor protein has been correlated with expression in a specific tissue, antibodies that are specific for this receptor protein can be used to identify a tissue type.

25       The antibodies are also useful in forensic identification. Accordingly, where an individual has been correlated with a specific genetic polymorphism resulting in a specific polymorphic protein, an antibody specific for the polymorphic protein can be used as an aid in identification.

30       The antibodies are also useful for inhibiting receptor function, for example, blocking ligand binding.

      These uses can also be applied in a therapeutic context in which treatment involves inhibiting receptor function. An antibody can be used, for example, to block ligand binding. Antibodies can be prepared against specific fragments

containing sites required for function or against intact receptor associated with a cell.

The invention also encompasses kits for using antibodies to detect the presence of a receptor protein in a biological sample. The kit can comprise  
5 antibodies such as a labeled or labelable antibody and a compound or agent for detecting receptor protein in a biological sample; means for determining the amount of receptor protein in the sample; and means for comparing the amount of receptor protein in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for  
10 using the kit to detect receptor protein.

#### Polynucleotides

The specifically disclosed cDNA comprises the coding region and 5' and 3' untranslated sequences (SEQ ID NO 2).

15 The human 14926 receptor cDNA is approximately 2818 nucleotides in length and encodes a full length protein that is approximately 370 amino acid residues in length. The nucleic acid is expressed in brain. Structural analysis of the amino acid sequence of SEQ ID NO 1 is provided in Figure 3, a hydropathy plot. The figure shows the putative structure of the seven transmembrane  
20 segments, the amino terminal extracellular domain and the carboxy terminal intracellular domain.

As used herein, the term "transmembrane segment" refers to a structural amino acid motif which includes a hydrophobic helix that spans the plasma membrane. The entire transmembrane domain spans from about amino acid 24 to  
25 about amino acid 341. Seven segments span the membrane and there are three intracellular and three extracellular loops in this domain.

The invention provides isolated polynucleotides encoding a 14926 receptor protein. The term "14926 polynucleotide" or "14926 nucleic acid" refers to the sequence shown in SEQ ID NO 2. The term "receptor polynucleotide" or "receptor  
30 nucleic acid" further includes variants and fragments of the 14926 polynucleotide.

An "isolated" receptor nucleic acid is one that is separated from other nucleic acid present in the natural source of the receptor nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic



acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB. The important point is that the nucleic acid is isolated from flanking sequences  
5 such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the receptor nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when  
10 produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include  
15 recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

20 The receptor polynucleotides can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten  
25 protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

The receptor polynucleotides include, but are not limited to, the sequence encoding the mature polypeptide alone, the sequence encoding the mature  
30 polypeptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature polypeptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3'

sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the polynucleotide may be fused to a marker sequence encoding, for example, a peptide that facilitates  
5 purification.

Receptor polynucleotides can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-  
10 stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

One receptor nucleic acid comprises the nucleotide sequence shown in SEQ ID NO 2, corresponding to human brain cDNA.

In one embodiment, the receptor nucleic acid comprises only the coding  
15 region.

The invention further provides variant receptor polynucleotides, and fragments thereof, that differ from the nucleotide sequence shown in SEQ ID NO 2 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence shown in SEQ ID NO 2.

20 The invention also provides receptor nucleic acid molecules encoding the variant polypeptides described herein. Such polynucleotides may be naturally occurring, such as allelic variants (same locus), homologs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be  
25 made by mutagenesis techniques, including those applied to polynucleotides, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions.

Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid  
30 substitutions.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. These variants comprise a nucleotide sequence encoding a receptor that is at least about 55-60%, 60-65%, 65-70%, typically at least about 70-

75%, more typically at least about 80-85%, and most typically at least about 90-95% or more homologous to the nucleotide sequence shown in SEQ ID NO 2 or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under stringent conditions, to the nucleotide sequence shown in SEQ ID NO 2 or a fragment of the sequence. It is understood that stringent hybridization does not indicate substantial homology where it is due to general homology, such as poly A sequences, or sequences common to all or most proteins, all GPCRs, or all family I GPCRs. Moreover, it is understood that variants do not include any of the nucleic acid sequences that may have been disclosed prior to the invention.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a receptor at least 55-60% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 65%, at least about 70%, or at least about 75% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C. In one embodiment, an isolated receptor nucleic acid molecule that hybridizes under stringent conditions to the sequence of SEQ ID NO 2 corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

Furthermore, the invention provides polynucleotides that comprise a fragment of the full length receptor polynucleotides. The fragment can be single or double stranded and can comprise DNA or RNA. The fragment can be derived from either the coding or the non-coding sequence.

In one embodiment, an isolated receptor nucleic acid is at least 404 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO 2. In another embodiment an isolated receptor nucleic acid encodes the entire coding region

from amino acid 1 to amino acid 370. In another embodiment the isolated receptor nucleic acid encodes a sequence corresponding to the mature protein from about amino acid 6 to amino acid 370. Fragments further include nucleic acid sequences encoding a portion of the amino acid sequence described herein and further including flanking nucleotide sequences at the 3' region. Other fragments include nucleotide sequences encoding the amino acid fragments described herein. Receptor nucleic acid fragments also include a fragment from nucleotide 1 to around nucleotide 43 and subfragments thereof; from about 1321 to about 1391 and subfragments thereof greater than 15 nucleotides; about 1391 to about 1526 and subfragments thereof; about 1527 to about 1565 and subfragments thereof greater than 14 nucleotides; about 1566 to about 1712 and subfragments thereof; and about 1800 to 2818 and subfragments thereof. In these embodiments, the nucleic acid can be at least 17, 20, 30, 40, 50, 100, 250, or 500 nucleotides in length or greater. Nucleic acid fragments, according to the present invention, are not to be construed as encompassing those fragments that may have been disclosed prior to the invention.

Receptor nucleic acid fragments further include sequences corresponding to the domains described herein, subregions also described, and specific functional sites. Receptor nucleic acid fragments include nucleic acid molecules encoding a polypeptide comprising the amino terminal extracellular domain including amino acid residues from 1 to about 23, a polypeptide comprising the region spanning the transmembrane domain (amino acid residues from about 24 to about 341), a polypeptide comprising the carboxy terminal intracellular domain (amino acid residues from about 342 to about 370), and a polypeptide encoding the G-protein receptor signature (118-120 or surrounding amino acid residues from about 109 to about 125), nucleic acid molecules encoding any of the seven transmembrane segments, extracellular or intracellular loops, glycosylation sites, cAMP or cGMP phosphorylation sites, protein kinase C phosphorylation sites and casein kinase II phosphorylation sites, myristoylation sites, and amidation site. Receptor nucleic acid fragments also include combinations of the domains, segments, loops, and other functional sites described above. Thus, for example, a receptor nucleic acid could include sequences corresponding to the amino terminal extracellular domain and one transmembrane fragment. A person of ordinary skill in the art would be

aware of the many permutations that are possible. Where the location of the domains have been predicted by computer analysis, one of ordinary skill would appreciate that the amino acid residues constituting these domains can vary depending on the criteria used to define the domains.

5 In one embodiment, an isolated receptor nucleic acid is at least 404 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO 2. In other embodiments, the nucleic acid is at least 40, 50, 100, 250 or 500 nucleotides in length. However, it is understood that a receptor fragment includes any  
10 nucleic acid sequence that does not include the entire gene.

The invention also provides receptor nucleic acid fragments that encode epitope bearing regions of the receptor proteins described herein.

The isolated receptor polynucleotide sequences, and especially fragments, are useful as DNA probes and primers.

15 For example, the coding region of a receptor gene can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of receptor genes.

20 A probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 7-12, typically about 25, more typically about 40, 50 or 75 consecutive nucleotides of SEQ ID NO 2 sense or anti-sense strand or other receptor polynucleotides. A probe further comprises a  
25 label, e.g., radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

#### Polynucleotide Uses

30 The receptor polynucleotides are useful for probes, primers, and in biological assays. Where the polynucleotides are used to assess GPCR properties or functions, such as in the assays described herein, all or less than all of the entire cDNA can be useful. In this case, even fragments that may have been known prior to the invention are encompassed. Thus, for example, assays specifically directed

to GPCR functions, such as assessing agonist or antagonist activity, encompass the use of known fragments. Further, diagnostic methods for assessing receptor function can also be practiced with any fragment, including those fragments that may have been known prior to the invention. Similarly, in methods involving  
5 treatment of receptor dysfunction, all fragments are encompassed including those which may have been known in the art.

The receptor polynucleotides are useful as a hybridization probe for cDNA and genomic DNA to isolate a full-length cDNA and genomic clones encoding the polypeptide described in SEQ ID NO 1 and to isolate cDNA and genomic clones  
10 that correspond to variants producing the same polypeptide shown in SEQ ID NO 1 or the other variants described herein. Variants can be isolated from the same tissue and organism from which the polypeptide shown in SEQ ID NO 1 was isolated, different tissues from the same organism, or from different organisms. This method is useful for isolating genes and cDNA that are developmentally-  
15 controlled and therefore may be expressed in the same tissue or different tissues at different points in the development of an organism.

The probe can correspond to any sequence along the entire length of the gene encoding the receptor. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as described  
20 herein, probes would not encompass fragments and sequences that may have been disclosed prior to the invention.

The nucleic acid probe can be, for example, the full-length cDNA of SEQ ID NO 1, or a fragment thereof, such as an oligonucleotide of at least 7, 10, 12, 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically  
25 hybridize under stringent conditions to mRNA or DNA.

Fragments of the polynucleotides described herein are also useful to synthesize larger fragments or full-length polynucleotides described herein. For example, a fragment can be hybridized to any portion of an mRNA and a larger or full-length cDNA can be produced.

30 The fragments are also useful to synthesize antisense molecules of desired length and sequence.

The receptor polynucleotides are also useful as primers for PCR to amplify any given region of a receptor polynucleotide.

The receptor polynucleotides are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the receptor polypeptides. Vectors also include insertion vectors, used to integrate into another polynucleotide sequence, such as into the cellular genome, to alter *in situ* expression of receptor genes and gene products. For example, an endogenous receptor coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations. The receptor polynucleotides are also useful as probes for determining the chromosomal positions of the receptor polynucleotides by means of *in situ* hybridization methods.

The receptor polynucleotide probes are also useful to determine patterns of the presence of the gene encoding the receptors and their variants with respect to tissue distribution, for example, whether gene duplication has occurred and whether the duplication occurs in all or only a subset of tissues. The genes can be naturally occurring or can have been introduced into a cell, tissue, or organism exogenously. The receptor polynucleotides are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from genes encoding the polynucleotides described herein.

The receptor polynucleotides are also useful for constructing host cells expressing a part, or all, of the receptor polynucleotides and polypeptides.

The receptor polynucleotides are also useful for constructing transgenic animals expressing all, or a part, of the receptor polynucleotides and polypeptides.

The receptor polynucleotides are also useful for making vectors that express part, or all, of the receptor polypeptides.

The receptor polynucleotides are also useful as hybridization probes for determining the level of receptor nucleic acid expression. Accordingly, the probes can be used to detect the presence of, or to determine levels of, receptor nucleic acid in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the polypeptides described herein can be used to assess gene copy number in a given cell, tissue, or organism. This is particularly relevant in cases in which there has been an amplification of the receptor genes.

Alternatively, the probe can be used in an *in situ* hybridization context to assess the position of extra copies of the receptor genes, as on extrachromosomal elements or as integrated into chromosomes in which the receptor gene is not normally found, for example as a homogeneously staining region.

5        These uses are relevant for diagnosis of disorders involving an increase or decrease in receptor expression relative to normal results.

*In vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

10       Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a receptor protein, such as by measuring a level of a receptor-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a receptor gene has been mutated.

15       Nucleic acid expression assays are useful for drug screening to identify compounds that modulate receptor nucleic acid expression.

20       The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the receptor gene. The method typically includes assaying the ability of the compound to modulate the expression of the receptor nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired receptor nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the receptor nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

25       Alternatively, candidate compounds can be assayed *in vivo* in patients or in transgenic animals.

30       The assay for receptor nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway (such as cyclic AMP or phosphatidylinositol turnover). Further, the expression of genes that are up- or down-regulated in response to the receptor protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.



Thus, modulators of receptor gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of receptor mRNA in the presence of the candidate compound is compared to the level of expression of receptor mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression. Accordingly, the invention provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate receptor nucleic acid expression. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) of nucleic acid expression.

Alternatively, a modulator for receptor nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the receptor nucleic acid expression.

The receptor polynucleotides are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the receptor gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The receptor polynucleotides are also useful in diagnostic assays for qualitative changes in receptor nucleic acid, and particularly in qualitative changes

that lead to pathology. The polynucleotides can be used to detect mutations in receptor genes and gene expression products such as mRNA. The polynucleotides can be used as hybridization probes to detect naturally-occurring genetic mutations in the receptor gene and thereby to determine whether a subject with the mutation  
5 is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the receptor gene associated with a  
10 dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a receptor protein.

Individuals carrying mutations in the receptor gene can be detected at the nucleic acid level by a variety of techniques. Genomic DNA can be analyzed  
15 directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way.

In certain embodiments, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in  
20 a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic,  
25 mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample.  
30 Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a receptor gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be  
5 used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

10 Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method.

Furthermore, sequence differences between a mutant receptor gene and a wild-type gene can be determined by direct DNA sequencing. A variety of  
15 automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

20 Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared  
25 (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point  
30 mutations include, selective oligonucleotide hybridization, selective amplification, and selective primer extension. The receptor polynucleotides are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the polynucleotides can be used

to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). In the present case, for example, a mutation in the receptor gene that results in altered affinity for ligand could result in an excessive or decreased drug effect with standard concentrations of ligand that activates the receptor. Accordingly, the receptor polynucleotides described herein can be used to assess the mutation content of the receptor gene in an individual in order to select an appropriate compound or dosage regimen for treatment.

Thus polynucleotides displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The receptor polynucleotides are also useful for chromosome identification when the sequence is identified with an individual chromosome and to a particular location on the chromosome. First, the DNA sequence is matched to the chromosome by *in situ* or other chromosome-specific hybridization. Sequences can also be correlated to specific chromosomes by preparing PCR primers that can be used for PCR screening of somatic cell hybrids containing individual chromosomes from the desired species. Only hybrids containing the chromosome containing the gene homologous to the primer will yield an amplified fragment. Sublocalization can be achieved using chromosomal fragments. Other strategies include prescreening with labeled flow-sorted chromosomes and preselection by hybridization to chromosome-specific libraries. Further mapping strategies include fluorescence *in situ* hybridization which allows hybridization with probes shorter than those traditionally used. Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on the chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

The receptor polynucleotides can also be used to identify individuals from small biological samples. This can be done for example using restriction fragment-length polymorphism (RFLP) to identify an individual. Thus, the polynucleotides described herein are useful as DNA markers for RFLP (See U.S. Patent No. 5,272,057).

Furthermore, the receptor sequence can be used to provide an alternative technique which determines the actual DNA sequence of selected fragments in the genome of an individual. Thus, the receptor sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify DNA from an individual for subsequent sequencing.

Panels of corresponding DNA sequences from individuals prepared in this manner can provide unique individual identifications, as each individual will have a unique set of such DNA sequences. It is estimated that allelic variation in humans occurs with a frequency of about once per each 500 bases. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. The receptor sequences can be used to obtain such identification sequences from individuals and from tissue. The sequences represent unique fragments of the human genome. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes.

If a panel of reagents from the sequences is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

The receptor polynucleotides can also be used in forensic identification procedures. PCR technology can be used to amplify DNA sequences taken from very small biological samples, such as a single hair follicle, body fluids (eg. blood, saliva, or semen). The amplified sequence can then be compared to a standard allowing identification of the origin of the sample.

The receptor polynucleotides can thus be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which

can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As described above, actual base sequence information can be used for identification as an accurate alternative to patterns  
5 formed by restriction enzyme generated fragments. Sequences targeted to the noncoding region are particularly useful since greater polymorphism occurs in the noncoding regions, making it easier to differentiate individuals using this technique. Fragments are at least 7-12 bases.

The receptor polynucleotides can further be used to provide polynucleotide  
10 reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue. This is useful in cases in which a forensic pathologist is presented with a tissue of unknown origin. Panels of receptor probes can be used to identify tissue by species and/or by organ type.

In a similar fashion, these primers and probes can be used to screen tissue  
15 culture for contamination (i.e. screen for the presence of a mixture of different types of cells in a culture).

Alternatively, the receptor polynucleotides can be used directly to block transcription or translation of receptor gene sequences by means of antisense or ribozyme constructs. Thus, in a disorder characterized by abnormally high or  
20 undesirable receptor gene expression, nucleic acids can be directly used for treatment. The receptor polynucleotides are thus useful as antisense constructs to control receptor gene expression in cells, tissues, and organisms. A DNA antisense polynucleotide is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of receptor  
25 protein. An antisense RNA or DNA polynucleotide would hybridize to the mRNA and thus block translation of mRNA into receptor protein.

Examples of antisense molecules useful to inhibit nucleic acid expression include antisense molecules complementary to a fragment of the 5' untranslated region of SEQ ID NO 2 which also includes the start codon and antisense  
30 molecules which are complementary to a fragment of the 3' untranslated region of SEQ ID NO 2.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of receptor nucleic acid. Accordingly,

these molecules can treat a disorder characterized by abnormal or undesired receptor nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the receptor protein, such as ligand binding.

The receptor polynucleotides also provide vectors for gene therapy in patients containing cells that are aberrant in receptor gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered ex vivo and returned to the patient, are introduced into an individual where the cells produce the desired receptor protein to treat the individual.

The invention also encompasses kits for detecting the presence of a receptor nucleic acid in a biological sample. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting receptor nucleic acid in a biological sample; means for determining the amount of receptor nucleic acid in the sample; and means for comparing the amount of receptor nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect receptor mRNA or DNA.

#### Vectors/host cells

The invention also provides vectors containing the receptor polynucleotides. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, that can transport the receptor polynucleotides. When the vector is a nucleic acid molecule, the receptor polynucleotides are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the receptor polynucleotides. Alternatively, the vector may integrate into the host cell genome

and produce additional copies of the receptor polynucleotides when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the receptor polynucleotides. The  
5 vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the receptor polynucleotides such that transcription of the polynucleotides is allowed in a host cell. The polynucleotides can be introduced into the host cell with a separate polynucleotide capable of affecting transcription.  
10 Thus, the second polynucleotide may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the receptor polynucleotides from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself.

15 It is understood, however, that in some embodiments, transcription and/or translation of the receptor polynucleotides can occur in a cell-free system.

The regulatory sequence to which the polynucleotides described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage  $\lambda$ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40,  
20 the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding  
25 sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription  
30 termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression



vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

- 5 A variety of expression vectors can be used to express a receptor polynucleotide. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses, papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses.
- 10 Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, eg. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring
- 15 Harbor, NY, (1989).

- The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive
- 20 and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

- The receptor polynucleotides can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the
- 25 expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

- The vector containing the appropriate polynucleotide can be introduced into an appropriate host cell for propagation or expression using well-known
- 30 techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the polypeptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the receptor polypeptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired polypeptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the polynucleotide of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

The receptor polynucleotides can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The receptor polynucleotides can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

In certain embodiments of the invention, the polynucleotides described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

5       The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the receptor polynucleotides. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the polynucleotides described herein. These are found for example  
10 in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but  
15 operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the polynucleotide sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences,  
20 constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

25       The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and  
30 other techniques such as those found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the receptor polynucleotides can be introduced either alone or with other polynucleotides that are not related to the receptor polynucleotides such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the receptor polynucleotide vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the polynucleotides described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the polypeptide is desired, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the receptor polypeptides or heterologous to these polypeptides.

Where the polypeptide is not secreted into the medium, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The polypeptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography,

hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the polypeptides described herein, the polypeptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the polypeptides may include an initial modified methionine in some cases as a result of a host-mediated process.

#### Uses of vectors and host cells

10 The host cells expressing the polypeptides described herein, and particularly recombinant host cells, have a variety of uses. First, the cells are useful for producing receptor proteins or polypeptides that can be further purified to produce desired amounts of receptor protein or fragments. Thus, host cells containing expression vectors are useful for polypeptide production.

15 Host cells are also useful for conducting cell-based assays involving the receptor or receptor fragments. Thus, a recombinant host cell expressing a native receptor is useful to assay for compounds that stimulate or inhibit receptor function. This includes ligand binding, gene expression at the level of transcription or translation, G-protein interaction, and components of the signal transduction pathway. Host cells are also useful for identifying receptor mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant receptor (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native receptor.

25 Recombinant host cells are also useful for expressing the chimeric polypeptides described herein to assess compounds that activate or suppress activation by means of a heterologous amino terminal extracellular domain (or other binding region). Alternatively, a heterologous region spanning the entire transmembrane domain (or parts thereof) can be used to assess the effect of a desired amino terminal extracellular domain (or other binding region) on any given host cell. In this embodiment, a region spanning the entire transmembrane domain (or parts thereof) compatible with the specific host cell is used to make the

chimeric vector. Alternatively, a heterologous carboxy terminal intracellular, e.g., signal transduction, domain can be introduced into the host cell.

Further, mutant receptors can be designed in which one or more of the various functions is engineered to be increased or decreased (e.g., ligand binding or G-protein binding) and used to augment or replace receptor proteins in an individual. Thus, host cells can provide a therapeutic benefit by replacing an aberrant receptor or providing an aberrant receptor that provides a therapeutic result. In one embodiment, the cells provide receptors that are abnormally active.

In another embodiment, the cells provide receptors that are abnormally inactive. These receptors can compete with endogenous receptors in the individual. In another embodiment, cells expressing receptors that cannot be activated, are introduced into an individual in order to compete with endogenous receptors for ligand. For example, in the case in which excessive ligand is part of a treatment modality, it may be necessary to inactivate this ligand at a specific point in treatment. Providing cells that compete for the ligand, but which cannot be affected by receptor activation would be beneficial.

Homologously recombinant host cells can also be produced that allow the *in situ* alteration of endogenous receptor polynucleotide sequences in a host cell genome. This technology is more fully described in WO 93/09222, WO 91/12650 and U.S. 5,641,670. Briefly, specific polynucleotide sequences corresponding to the receptor polynucleotides or sequences proximal or distal to a receptor gene are allowed to integrate into a host cell genome by homologous recombination where expression of the gene can be affected. In one embodiment, regulatory sequences are introduced that either increase or decrease expression of an endogenous sequence. Accordingly, a receptor protein can be produced in a cell not normally producing it, or increased expression of receptor protein can result in a cell normally producing the protein at a specific level. Alternatively, the entire gene can be deleted. Still further, specific mutations can be introduced into any desired region of the gene to produce mutant receptor proteins. Such mutations could be introduced, for example, into the specific functional regions such as the ligand-binding site or the G-protein binding site.

In one embodiment, the host cell can be a fertilized oocyte or embryonic stem cell that can be used to produce a transgenic animal containing the altered

receptor gene. Alternatively, the host cell can be a stem cell or other early tissue precursor that gives rise to a specific subset of cells and can be used to produce transgenic tissues in an animal. See also Thomas *et al.*, *Cell* 51:503 (1987) for a description of homologous recombination vectors. The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous receptor gene is selected (see e.g., Li, E. *et al.*, *Cell* 69:915 (1992)). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see e.g., Bradley, A. in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, E.J. Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, A. (1991) *Current Opinion in Biotechnology* 2:823-829 and in PCT International Publication Nos. WO 90/11354; WO 91/01140; and WO 93/04169.

The genetically engineered host cells can be used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a receptor protein and identifying and evaluating modulators of receptor protein activity.

Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

In one embodiment, a host cell is a fertilized oocyte or an embryonic stem cell into which receptor polynucleotide sequences have been introduced.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any

of the receptor nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the receptor protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* *PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al.* *Science* 251:1351-1355 (1991). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.



Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G<sub>0</sub> phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the polypeptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, receptor activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* receptor function, including ligand interaction, the effect of specific mutant receptors on receptor function and ligand interaction, and the effect of chimeric receptors. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more receptor functions.

In general, methods for producing transgenic animals include introducing a nucleic acid sequence according to the present invention, the nucleic acid sequence capable of expressing the receptor protein in a transgenic animal, into a cell in culture or *in vivo*. When introduced *in vivo*, the nucleic acid is introduced into an intact organism such that one or more cell types and, accordingly, one or more tissue types, express the nucleic acid encoding the receptor protein. Alternatively, the nucleic acid can be introduced into virtually all cells in an organism by transfecting a cell in culture, such as an embryonic stem cell, as described herein for the production of transgenic animals, and this cell can be used to produce an entire transgenic organism. As described, in a further embodiment, the host cell can be a fertilized oocyte. Such cells are then allowed to develop in a female foster animal to produce the transgenic organism.

Pharmaceutical compositions

5 The receptor nucleic acid molecules, protein (particularly fragments such as the amino terminal extracellular domain), modulators of the protein, and antibodies (also referred to herein as "active compounds") can be incorporated into pharmaceutical compositions suitable for administration to a subject, e.g., a human. Such compositions typically comprise the nucleic acid molecule, protein, modulator, or antibody and a pharmaceutically acceptable carrier.

10 As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, such  
15 media can be used in the compositions of the invention. Supplementary active compounds can also be incorporated into the compositions. A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal  
20 (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such  
25 as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable  
30 syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For

intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a receptor protein or anti-receptor antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For oral administration, the agent can be contained in enteric forms to survive the stomach or further coated or mixed to be released in a particular region of the GI tract by

known methods. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to

those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be  
5 prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for  
10 the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the  
15 limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. 5,328,470) or by  
20 stereotactic injection (see e.g., Chen *et al.*, *PNAS* 91:3054-3057 (1994)). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral  
25 vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

This invention may be embodied in many different forms and should not be  
30 construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will fully convey the invention to those skilled in the art. Many modifications and other embodiments of the invention will come to mind in one skilled in the art to which this invention

pertains having the benefit of the teachings presented in the foregoing description. Although specific terms are employed, they are used as in the art unless otherwise indicated.

## THAT WHICH IS CLAIMED:

1. An isolated polypeptide having an amino acid sequence selected from the group consisting of:
  - 5 (a) The amino acid sequence shown in SEQ ID NO 1;
  - (b) The amino acid sequence of an allelic variant of the amino acid sequence shown in SEQ ID NO 1;
  - (c) The amino acid sequence of a sequence variant of the amino acid sequence shown in SEQ ID NO 1, wherein the sequence variant is encoded by a  
10 nucleic acid molecule hybridizing to the nucleic acid molecule shown in SEQ ID NO 2 under stringent conditions;
  - (d) The amino acid sequence of the mature receptor polypeptide from about amino acid 6 to amino acid 370, shown in SEQ ID NO 1;
  - (e) The amino acid sequence of the polypeptide shown in SEQ ID NO  
15 1, from about amino acid 1 to about amino acid 23; and
  - (f) The amino acid sequence of an epitope bearing region of any one of the polypeptides of (a)-(f).
2. An isolated antibody that selectively binds to a polypeptide of claim 1, (a)-  
20 (f).
3. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:
  - (a) The nucleotide sequence shown in SEQ ID NO 2;
  - 25 (b) A nucleotide sequence encoding the amino acid sequence shown in SEQ ID NO 1; and
  - (c) A nucleotide sequence complementary to either of the nucleotide sequences in (a) or (b).
- 30 4. An isolated nucleic acid molecule having a nucleotide sequence selected from the group consisting of:
  - (a) A nucleotide sequence encoding an amino acid sequence of a sequence variant of the amino acid sequence shown in SEQ ID NO 1 that

hybridizes to the nucleotide sequence shown in SEQ ID NO 2 under stringent conditions; and

(b) A nucleotide sequence complementary to the nucleotide sequence in (a).

5

5. A nucleic acid vector comprising the nucleic acid sequences in any of claims 3-4.

6. A host cell containing the vector of claim 5.

10

7. A method for producing any of the polypeptides in claim 1 comprising introducing a nucleotide sequence encoding any of the polypeptide sequences in (a)-(f) into a host cell, and culturing the host cell under conditions in which the proteins are expressed from the nucleic acid.

15

8. A method for detecting the presence of any of the polypeptides in claim 1 in a sample, said method comprising contacting said sample with an agent that specifically allows detection of the presence of the polypeptide in the sample and then detecting the presence of the polypeptide.

20

9. The method of claim 8, wherein said agent is capable of selective physical association with said polypeptide.

10. The method of claim 9, wherein said agent binds to said polypeptide.

25

11. The method of claim 10, wherein said agent is an antibody.

12. The method of claim 10, wherein said agent is a ligand.

30 13. A kit comprising reagents used for the method of claim 8, wherein the reagents comprise an agent that specifically binds to said polypeptide.



14. A method for detecting the presence of any of the nucleic acid sequences in any of claims 3-4 in a sample, the method comprising contacting the sample with an oligonucleotide that hybridizes to the nucleic acid sequences under stringent conditions and determining whether the oligonucleotide binds to the nucleic acid sequence in the sample.
15. The method of claim 14, wherein the nucleic acid, whose presence is detected, is mRNA.
16. A kit comprising reagents used for the method of claim 14, wherein the reagents comprise a compound that hybridizes under stringent conditions to any of the nucleic acid molecules.
17. A method for identifying an agent that binds to any of the polypeptides in claim 1, said method comprising contacting the polypeptide with an agent that binds to the polypeptide and assaying the complex formed with the agent bound to the polypeptide.
18. The method of claim 17 wherein a fragment of the polypeptide is contacted.
19. A method for identifying an agent that modulates the level or activity of any of the polypeptides in claim 1 in a cell, the method comprising contacting the agent with a cell capable of expressing said polypeptide such that said polypeptide level or activity can be modulated in said cell by said agent and measuring said polypeptide level or activity.
20. A method for identifying an agent that interacts with any of the polypeptides in claim 1 in a cell, the method comprising contacting said agent with a cell capable of allowing an interaction between said polypeptide and said agent such that said polypeptide can interact with said agent and measuring the interaction.

21. A method of screening a cell to identify an agent that modulates the level or activity of any of the polypeptides in claim 1 in said cell, said method comprising contacting said agent with a cell capable of expressing said polypeptide such that said polypeptide level or activity can be modulated in said cell by said agent and  
5 measuring said polypeptide level or activity.

22. A method of screening a cell to identify an agent that interacts with any of the polypeptides in claim 1, said method comprising contacting said agent with a cell capable of allowing an interaction between said polypeptide and said agent  
10 such that said polypeptide can interact with said agent and measuring the interaction.

23. A method for modulating the activity of any of the polypeptides in claim 1, the method comprising contacting any of the polypeptides of claim 1 with an agent  
15 under conditions that allow the agent to modulate the activity of the polypeptide.

24. The method of claim 23 wherein said modulation is in cells derived from tissues selected from the group consisting of brain, spleen, lung, kidney, skeletal muscle, liver and heart.  
20

25. The method of claim 24 wherein said cells are brain cells.

26. The method of claim 23 wherein said modulation is in a patient having a disorder involving hyperplasia or inflammation.  
25

27. A pharmaceutical composition containing any of the polypeptides in claim 1 in a pharmaceutically acceptable carrier.

28. A pharmaceutical composition containing any of the nucleic acid molecules of claims 3-5 in a pharmaceutically acceptable carrier.  
30

29. A non human transgenic animal wherein one or more cells of said animal contains any of the nucleic acid sequences of claims 3-5.

30. A non human transgenic animal wherein one or more cells of said animal contains any of the nucleic acid sequences of claims 3-5 wherein said cell expresses any of the polypeptides of claim 1.

5

31. A method for producing a transgenic animal according to claim 30, said method comprising introducing any of the nucleic acid sequences of claim 1 into a cell, wherein said cell is present in said animal or gives rise to said animal.

1/6

Input file flh14926cons; Output File flh14926tra  
Sequence length 2818

CCACGCGTCCGATTATAGAAAGAAGCTATAGTCTGTGCGGATATCCAACACCAACAGGGCTTACTGAGAGCTCCATTT  
CTGGAAAGCCTTACAAGACTGAGGAATATCAGACTGCGAATCACCAGGAACGGTTCCTTGACAGACAGAAGCAATCTC  
TCTCCCCATCTTCGCATATTCTAATGGCAAAACAAGTGGGAAGAAAGAGGAAGCATGACTGCAGATCAGATCAGTTCTC  
TTTGTGGATTATATTTTCAGTAAATGTATGGATCTATCTTTTCCTTGTCTTATATCTAGATCATGAGACTTGACTGA

	M A N Y S H A A D N I L Q	13
GGCTGTATCCTTATCCTCCATCCATCT	ATG GCG AAC TAT AGC CAT GCA GCT GAC AAC ATT TTG CAA	39
N L S P L T A F L K L T S L G F I I G V		33
AAT CTC TCG CCT CTA ACA GCC TTT CTG AAA CTG ACT TCC TTG GGT TTC ATA ATA GGA GTC		99
S V V G N L L I S I L L V K D K T L H R		53
AGC GTG GTG GGC AAC CTC CTG ATC TCC ATT TTG CTA GTG AAA GAT AAG ACC TTG CAT AGA		159
A P Y Y F L L D L C C S D I L R S A I C		73
GCA CCT TAC TAC TTC CTG TTG GAT CTT TGC TGT TCA GAT ATC CTC AGA TCT GCA TTT TGT		219
F P F V F N S V K N G S T V T Y G T L T		93
TTC CCA TTT GTG TTC AAC TCT GTC AAA AAT GGC TCT ACC TGG ACT TAT GGG ACT CTG ACT		279
C K V I A F L G V L S C F H T A F M L F		113
TGC AAA GTG ATT GCC TTT CTG GGG GTT TTG TCC TGT TTC CAC ACT GCT TTC ATG CTC TTC		339
C I S V T R Y L A I A H H R F Y T K R L		133
TGC ATC AGT GTC <u>ACC AGA TAC</u> TTA GCT ATC GCC CAT CAC CGC TTC TAT ACA AAG AGG CTG		399
T F V T C L A V I C M V W T L S V A M A		153
ACC TTT TGG ACG TGT CTG GCT GTG ATC TGT ATG GTG TGG ACT CTG TCT GTG GCC ATG GCA		459
F P P V L D V G T Y S F I R E E D Q C T		173
TTT CCC CCG GTT TTA GAC GTG GGC ACT TAC TCA TTC ATT AGG GAG GAA GAT CAA TGC ACC		519
F Q H R S F R A N D S L G F M L L L A L		193
TTC CAA CAC CGC TCC TTC AGG GCT AAT GAT TCC TTA GGA TTT ATG CTG CTT CTT GCT CTC		579
I L L A T Q L V Y L K L I F F V H D R R		213
ATC CTC CTA GCC ACA CAG CTT GTC TAC CTC AAG CTG ATA TTT TTC GTC CAC GAT CGA AGA		639
K M K P V Q F V A A V S Q N W T F H G P		233
AAA ATG AAG CCA GTC CAG TTT GTA GCA GCA GTC AGC CAG AAC TGG ACT TTT CAT GGT CCT		699
G A S G Q A A A N V L A G F G R G P T P		253
GGA GCC AGT GGC CAG GCA GCT GCC AAT TGG CTA GCA GGA TTT GGA AGG GGT CCC ACA CCA		759
P T L L G I R Q N A N T T G R R R L L V		273
CCC ACC TTG CTG GGC ATC AGG CAA AAT GCA AAC ACC ACA GGC AGA AGA AGG CTA TTG GTC		819
L D E F K M E K R I S R M F Y I M T F L		293
TTA GAC GAG TTC AAA ATG GAG AAA AGA ATC AGC AGA ATG TTC TAT ATA ATG ACT TTT CTG		879
F L T L V G P Y L V A C Y V R V F A R G		313
TTT CTA ACC TTG TGG GGC CCC TAC CTG GTG GCC TGT TAT TGG AGA GTT TTT GCA AGA GGG		939

TO FIG. 1B.

2/6

FROM FIG. 1A.

P	V	V	P	G	G	F	L	T	A	A	V	V	M	S	F	A	Q	A	G	333
CCT	GTA	GTA	CCA	GGG	GGA	TTT	CTA	ACA	GCT	GCT	GTC	TGG	ATG	AGT	TTT	GCC	CAA	GCA	GGA	999
I	N	P	F	V	C	I	F	S	N	R	E	L	R	R	C	F	S	T	T	353
TTT	CCC	CCG	GTT	TTA	GAC	GTG	GGC	ACT	TAC	TCA	TTC	ATT	AGG	GAG	GAA	GAT	CAA	TGC	ACC	1059
L	L	Y	C	R	K	S	R	L	P	R	E	P	Y	C	V	I	*			371
CTT	CTT	TAC	TGC	AGA	AAA	TCC	AGG	TTA	CCA	AGG	GAA	CCT	TAC	TGT	GTT	ATA	TGA			1113

GGGAGCATCTGTAATCTTTAGCCTTGTGAAACTAACCTTCTCTGCTGAGCAATTGTGGCCCATAGCCATATTTTGAG  
 AAGGAAATTCAGAATGGAATCAGCAGTTTTAAGGATTTGGGCAACATTCTGCAGTCTTTGCAATAGTTCACCTATAATC  
 CTATTTTAAATCTCAGAGTGATCCTGCTGACTGCCAGCAAAGGTTTGTAAATTAAGAAGGGACTGAACCACTGCCCTAAG  
 TTTCTTTATGTGGTCAAAACTAGATAATGAAAGTAGCAGGTGCTAAGTATCAGTGCTAAATGCTCTGTATGCTACTAC  
 ATATGAAAAACATAAAAAACAATTAGCATTGGACATCTTAATAAATTAAGTTGACATGAGGTAATGTGTTGATAAA  
 AACTAATTTTAGAAGTTTGAAGACTTTAAACATTTCACTACTATTGTTTTGCAAAGACTAAATATTTGGGGACTT  
 AAAGTACTGTAATCCACTAAAGACGTGCCAATGAATTATTGGAATATCACACTTTAAAACCGCCTTGTAAGTTCTGGG  
 GAGCATTCCAAAGCAGTATATTGGTTCCAATTAGAGTTACTTTTTTGTATTAATACATTGCTATTTCTAAATACCAC  
 TTTCTCATCTACTAGTAAGATTGCTAGCATTGAAGTGTATTATGTGGTTTTTGTGATTGGTATAAAGTTTTTCCAA  
 TTCATTTATATTTTACAAATGCTAGATATTGGTCTGGGAGGCAACATTAATGGTACCAGCCTGTCACAACTGAGCAGTT  
 CTAATAATGCAGAATAAACACATGTTGCCTTAAAGGGTTATCTAGKATCCYTTATCTTATTAGCACTGGAGCAAATAG  
 YCAAGGGAAATCRAATCAGTAACTGGTCATGGTCATGCATCTRAAGTGCATGGAAGATCATTTAGTACTTTTTCTTT  
 TTTCTCACATGGTTTGAACCTTAAAGTGACATCMCTGAAATAATGAGATTTTCTTTTMRGGTGTGCTACCCCTYYTAR  
 ASTGTTCTAAGAAGCAGGCAGTTGATGTATGTTTATTTTAAGTCAGCTGTCGAGGGGAGACCACAGCCCTTAGTATGA  
 CATCTGCACAATTTGTGAAGCATTTATTCTACTGAAGGCACAGTCTTGTTTATACTTTCTGCACATTCAGTGTATTG  
 TCATTTAAATTATTTAGTTTTAACTTGTAAGCTTATAATATGATTTCTGGTATTTTAGAAATACATTAGAGTCTGT  
 GAGTCTCATTCTTTAAGATACANATGTGTGAAGCTCAATATAAGTTGCATTTGCCAAAATTTAAAAAAAAAAAAAAAA  
 AAAAAAAAAA

FIG. 1B.

3/6

## Protein Family/Domain HMM Matches for f1h14926orfaa

&gt;PF00001/7th\_1 7 transmembrane receptor (rhodopsin family)

Score: 146.01 Seq: 37 339 Model: 1 269

f1h14926or 37 GNILVIWvIcRyRRMRTPMNYFIvNLAvADLLFs1ftMPFWMvYyvMqg  
 GN+L+ +++++ + +++ ++YF++ L +D+L S ++PF + + ++  
 GNLLISILLVKDKTLHRAPYYFLDLCCSDILRSAICFPFVNSVKNGS 85

f1h14926or 86 RWpFGdfMCRiWmYFDYMNMYASIFFLtCISIDRYLWAICHPMrYmRWMT  
 +W++G++ C+++ ++ +++++ F+L CIS+ RYL AI H Y + T  
 TWYGTLTCKVIAFLGLVLSCHFATMLFCISVIRYL-AIAHHRFYTKRLT 134

f1h14926or 135 pRHRAWvMIiIwvMSF1ISMPPFLMFr. WstyrDEneWmTWCMiYDWP  
 + +++++I+++W++S++++PP+L + +S+ R E++ C++ +  
 FW-TCLAVICMVWTLVSAMAFPPVLDVGTYSFIREEDQ-----CTF-Q-- 175

f1h14926or 176 ewMWvWYvILmtiimgFYIPMiIM1FCYwRIYRIaRIWMRMipswQr...  
 +R++ +GF++ + ++L ++Y + ++ + ++  
 ---HRSFR-ANDS-LGFMLLLALILLATQLVYLKLIFVHDR---RKMKP 217

f1h14926or 218 .....  
 VQFVAAVSQNWTFHGPASGQAAANWLAGFGRGPTPTLLGIRQNANTTG 267

f1h14926or 268 RRrms....mRrERRivKM1iIImvVFIIICWIPYFIvmfMDTLMmWwFCe  
 RRR+ +++E+Ri++M I+ ++F+ W+PY ++ + +++F  
 RRRLVLDEFKMEKRISRMFYIMTFLTLWGPYLVACY-----WRVFAR 312

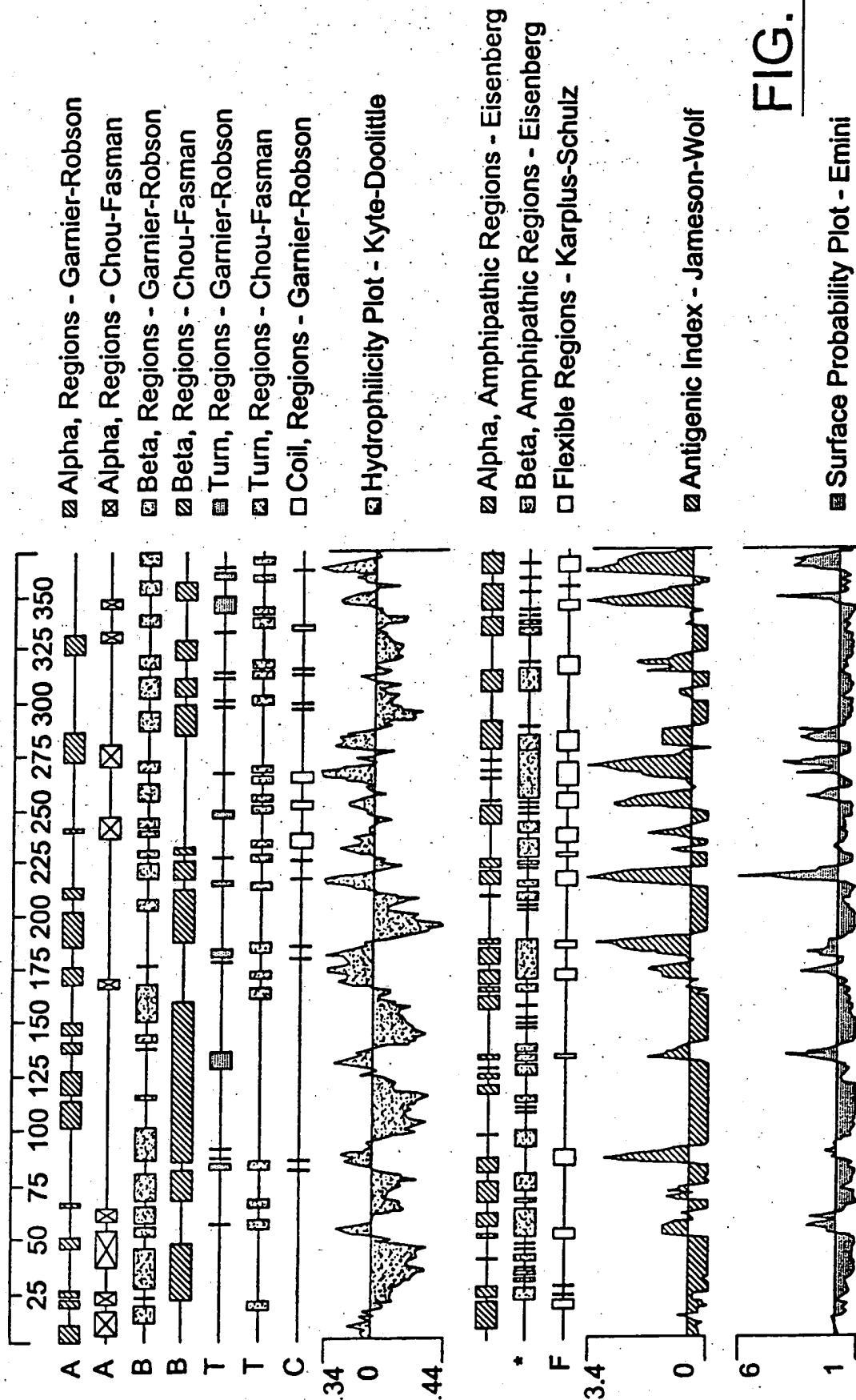
f1h14926or 313 fC.IwrrIWmYIfeWLaYvNCCpCiNPIIY\*  
 ++ + +++W++++ INP++  
 GPVVPGGFLT-AAVWMSFAQA-GINPFVC 339

&gt;MILPAT00028/ngf NGF / BDNF / Neurotrophins 3,4, and 6 family of cytokines

Score: 0.47 Seq: 290 302 Model: 1 13

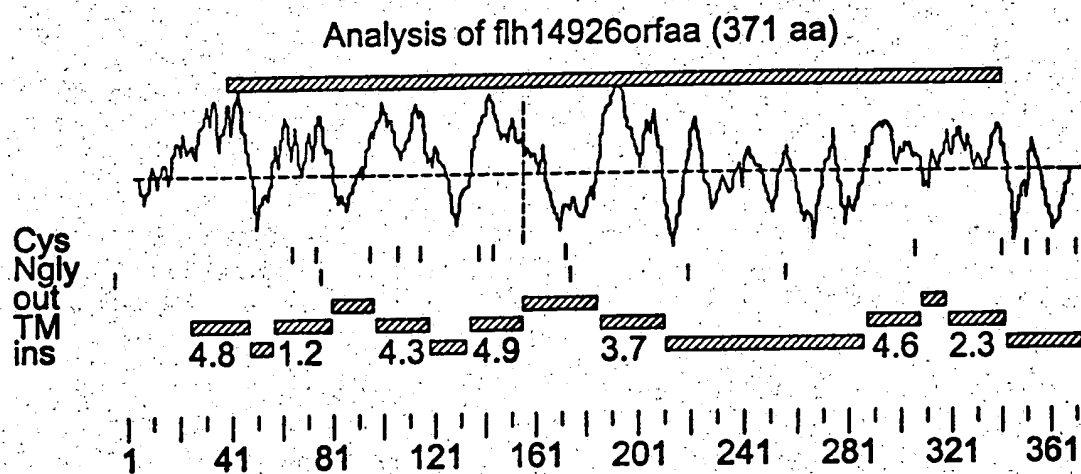
REF xxxxxxxxxxxxxxxx  
 \*MSMLFYTMFI<sub>s</sub>YF\*  
 M+ LF+T+ +Y+  
 f1h14926or 290 MTLFLTLWGPYL 302

**FIG. 2.**



25 50 75 100 125 150 175 200 225 250 275 300 325 350

5/6

FIG. 4.



6/6

# prositescan will scan one or more sequences against a set of sequence patterns

Database: Release 12.2 of February 1995

Tue Apr 7 18:49:19 1998

1025 patterns

Query=f114926orfaa

>PS00001/PDOC0001/ASN\_GLYCOSYLATION N-glycosylation site.  
N[\*P] [ST][\*P]

Query: 3 NYSH 6

Query: 83 NYSH 86

Query: 182 NYSH 185

Query: 227 NYSH 230

Query: 264 NYSH 267

>PS00004/PDOC0004/CAMP\_PHOSPHO\_SITE cAMP- and cGMP-dependent protein kinase  
phosphorylation site.  
[RD][2][A-Z][st]

Query: 131 KRLT 134

Query: 281 KRIS 284

>PS00005/PDOC0005/PDC\_PHOSPHO\_SITE Protein kinase C phosphorylation site.  
[ST][A-Z][RK]

Query: 80 SKV 82

Query: 93 TCK 95

Query: 130 TKR 132

Query: 178 SFR 180

Query: 266 TGR 268

Query: 342 SNR 344

>PS00006/PDOC0006/CK2\_PHOSPHO\_SITE Casein kinase II phosphorylation site.  
[ST][A-Z][2][DE]

Query: 342 SNRE 345

>PS00008/PDOC0008/MYRISTYL N-myristoylation site.  
G[\*EDRKHPFYW][A-Z][2][STAGCN][\*P]

Query: 84 GSTWTY 89

Query: 90 GTLTCK 95

Query: 101 GVLSCF 106

Query: 237 GQAAAN 242

Query: 258 GIRQNA 263

Query: 318 GGFLTA 323

>PS00009/PDOC0009/AMIDATION Amidation site.  
[A-Z]G[RK][RK]

Query: 266 TGRR 269

**FIG. 5.**

## SEQUENCE LISTING

&lt;110&gt; Glucksmann, Maria A.

&lt;120&gt; 14926 Receptor, A Novel G-Protein Coupled Receptor

&lt;130&gt; 035800/169197

&lt;140&gt; 09/145,745

&lt;141&gt; 1998-09-02

&lt;160&gt; 3

&lt;170&gt; PatentIn Ver. 2.0

&lt;210&gt; 1

&lt;211&gt; 370

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 1

Met Ala Asn Tyr Ser His Ala Ala Asp Asn Ile Leu Gln Asn Leu Ser  
1 5 10 15

Pro Leu Thr Ala Phe Leu Lys Leu Thr Ser Leu Gly Phe Ile Ile Gly  
20 25 30

Val Ser Val Val Gly Asn Leu Leu Ile Ser Ile Leu Leu Val Lys Asp  
35 40 45

Lys Thr Leu His Arg Ala Pro Tyr Tyr Phe Leu Leu Asp Leu Cys Cys  
50 55 60

Ser Asp Ile Leu Arg Ser Ala Ile Cys Phe Pro Phe Val Phe Asn Ser  
65 70 75 80

Val Lys Asn Gly Ser Thr Trp Thr Tyr Gly Thr Leu Thr Cys Lys Val  
85 90 95

Ile Ala Phe Leu Gly Val Leu Ser Cys Phe His Thr Ala Phe Met Leu  
100 105 110

Phe Cys Ile Ser Val Thr Arg Tyr Leu Ala Ile Ala His His Arg Phe  
115 120 125

Tyr Thr Lys Arg Leu Thr Phe Trp Thr Cys Leu Ala Val Ile Cys Met  
130 135 140

Val Trp Thr Leu Ser Val Ala Met Ala Phe Pro Pro Val Leu Asp Val  
145 150 155 160

Gly Thr Tyr Ser Phe Ile Arg Glu Glu Asp Gln Cys Thr Phe Gln His  
165 170 175

Arg Ser Phe Arg Ala Asn Asp Ser Leu Gly Phe Met Leu Leu Leu Ala  
180 185 190

Leu Ile Leu Leu Ala Thr Gln Leu Val Tyr Leu Lys Leu Ile Phe Phe  
195 200 205

Val His Asp Arg Arg Lys Met Lys Pro Val Gln Phe Val Ala Ala Val  
210 215 220

Ser Gln Asn Trp Thr Phe His Gly Pro Gly Ala Ser Gly Gln Ala Ala  
225 230 235 240

Ala Asn Trp Leu Ala Gly Phe Gly Arg Gly Pro Thr Pro Pro Thr Leu  
245 250 255

Leu Gly Ile Arg Gln Asn Ala Asn Thr Thr Gly Arg Arg Arg Leu Leu  
260 265 270

Val Leu Asp Glu Phe Lys Met Glu Lys Arg Ile Ser Arg Met Phe Tyr  
275 280 285

Ile Met Thr Phe Leu Phe Leu Thr Leu Trp Gly Pro Tyr Leu Val Ala  
290 295 300

Cys Tyr Trp Arg Val Phe Ala Arg Gly Pro Val Val Pro Gly Gly Phe  
305 310 315 320

Leu Thr Ala Ala Val Trp Met Ser Phe Ala Gln Ala Gly Ile Asn Pro  
325 330 335

Phe Val Cys Ile Phe Ser Asn Arg Glu Leu Arg Arg Cys Phe Ser Thr  
340 345 350

Thr Leu Leu Tyr Cys Arg Lys Ser Arg Leu Pro Arg Glu Pro Tyr Cys  
355 360 365

Val Il  
370

<210> 2  
<211> 2810

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 2

```
ccacgcgtcc gatttataga aagaagctat agtctgtcgg gatatccaac accaacaggg 60
cttactgaga gctccatttc tggaaagcct tacaagactg aggaatatca gactgcgaat 120
caccgggaac ggttcctttg cagcacagaa gcaatctctc tccccatctt cgcataattct 180
aatggcaaaa caagtggag aaagaggaa gcatgactgc agatcagatc agttctcttt 240
gtggattata ttttcagtaa aatgtatgga tctatctttt ccttggtctt atatctagat 300
catgagactt gactgaggct gtatccttat cctccatcca tctatggcga actatagcca 360
tgcagctgac aacattttgc aaaatctctc gcctctaaca gcctttctga aactgacttc 420
cttgggtttc ataataggag tcagcgtggt gggcaacctc ctgatctcca ttttgctagt 480
gaaagataag accttgcata gagcacctta ctacttctctg ttggatcttt gctgttcaga 540
tatctcaga tctgcaattt gtttccatt tgtgttcaac tctgtcaaaa atggctctac 600
ctggacttat gggactctga cttgcaaagt gattgccttt ctgggggttt tgtcctgttt 660
ccacactgct ttcattgctt tctgcatcag tgtcaccaga tacttagcta tcgcccata 720
ccgcttctat acaaagaggc tgaccttttg gacgtgtctg gctgtgatct gtatggtgtg 780
gactctgtct gtggccatgg catttcccc gggttttagac gtgggcactt actcattcat 840
tagggaggaa gatcaatgca cttccaaca ccgctccttc agggctaatt attccttagg 900
atztatgctg cttcttgctc tcactctcct agccacacag cttgtctacc tcaagctgat 960
attttctgct cacgatcgaa gaaaaatgaa gccagtccag tttgtagcag cagtcagcca 1020
gaactggact tttcatggc ctggagccag tggccaggca gctgccatt ggctagcagg 1080
atttggaagg ggtccacac caccacctt gctgggcac aggcaaaatg caaacaccac 1140
aggcagaaga aggctattgg tcttagacga gttcaaaatg gagaaaagaa tcagcagaat 1200
gttctatata atgacttttc tgtttctaac cttgtggggc ccctacctgg tggcctgtta 1260
ttggagagtt tttgcaagag ggctgtagt accaggggga tttctaacag ctgctgtctg 1320
gatgagtttt gcccaagcag gaatcaatcc tttgtctgc attttctcaa acagggagct 1380
gaggcgctgt ttcagcacia cccttcttta ctgcagaaaa tccaggttac caagggaacc 1440
ttactgtgtt atatgagggg gcatctgtaa atcttttagcc ttgtgaaaac taaccttctc 1500
tgctgagcaa ttgtggccca tagccatatt ttgagaagaa attcaagaat ggaatcagca 1560
gttttaagga tttgggcaac attctgcagt ctttgcaata gttcacctat aatcctattt 1620
taaattctag agtgatcctg ctgactgcca gcaaaggttt gtaattaaga agggactgaa 1680
ccactgccct aagtttcttt atgtggtcaa aaactagata atgaaagtag cagggtgctaa 1740
gtatcagtgc taaatgctct gtatgtcact acatatgaaa aaacatcaaa aaacaattag 1800
cattggacat cttataaat taagttgaca tgaggtaaat gtgttgataa aaactaattt 1860
tagaagtttg aagacttta aacatttcat actactattg ttttgcaaag actaaaatat 1920
ttggggactt aaagtactgt aatccactaa agacgtgcc atgaattatt ggaatatcac 1980
actttaaaaa ccgccttgta agttctgggg agcattccaa agcagtatat tggttccaat 2040
tagagtttac ttttttgta ttaatacatt gctatttcta aataccactt tctcatcta 2100
ctagtaagat tgctagcatt gaactgtatt atgtggtttt tgttgatttg gtataaagtt 2160
tttccaattc atttatattt tacaaatgct agatattggt ctgggaggca acattaatgg 2220
taccagcctg tcacaactga gcagttctaa taatgcagaa taaacacatg ttgccttaaa 2280
gggttatcta gkatccyttc atcttattag cactggagca aatagycaag ggaaatcraa 2340
tcagtaactg gtcattgtca tgcattctaa agtgcattga agatcattta gtactttttc 2400
cttttttctc acatggtttg aaacttaaag tgcacatcmc tgaaataatg agattttctt 2460
ttmgtgtg ctacccttyy tarastgttc taagaagcag gcagttgatg tatgtttata 2520
ttttaagtca gctgtcgagg ggagaccaca gccttagtat gacatcctgc acaatttgtg 2580
aagcatttat tctactgaag gcacagtctt gtttatactt tctgcacatt cagtgtattg 2640
```

gtcattttaa ttatttcagt ttttaacttgt gaaagcttat aatatgattt ctggtratttt 2700  
 agaaatacat tagagtctgt gagtctcatt ctttaagata canatgtgtg aacttcaata 2760  
 taaagttgca ttgccaataa ttttaaaaaa aaaaaaaaaa aaaaaaaaaa 2810

<210> 3

<211> 269

<212> PRT

<213> Unknown

<220>

<223> Description of Unknown Organism: Transmembrane  
 Receptor of the Rhodopsin Superfamily

<400> 3

Gly Asn Ile Leu Val Ile Trp Val Ile Cys Arg Tyr Arg Arg Met Arg  
 1 5 10 15

Thr Pro Met Asn Tyr Phe Ile Val Asn Leu Ala Val Ala Asp Leu Leu  
 20 25 30

Phe Ser Leu Phe Thr Met Pro Phe Trp Met Val Tyr Tyr Val Met Gly  
 35 40 45

Gly Arg Trp Pro Phe Gly Asp Phe Met Cys Arg Ile Trp Met Tyr Phe  
 50 55 60

Asp Tyr Met Asn Met Tyr Ala Ser Ile Phe Phe Leu Thr Cys Ile Ser  
 65 70 75 80

Ile Asp Arg Tyr Leu Trp Ala Ile Cys His Pro Met Arg Tyr Met Arg  
 85 90 95

Trp Met Thr Pro Arg His Arg Ala Trp Val Met Ile Ile Ile Ile Trp  
 100 105 110

Val Met Ser Phe Leu Ile Ser Met Pro Pro Phe Leu Met Phe Arg Trp  
 115 120 125

Ser Thr Tyr Arg Asp Glu Asn Glu Trp Asn Met Thr Trp Cys Met Ile  
 130 135 140

Tyr Asp Trp Pro Glu Trp Met Trp Arg Trp Tyr Val Ile Leu Met Thr  
 145 150 155 160

Ile Ile Met Gly Phe Tyr Ile Pro Met Ile Ile Met Leu Phe Cys Tyr  
 165 170 175

Trp Arg Ile Tyr Arg Ile Ala Arg Leu Trp Met Arg Met Ile Pro Ser



# INTERNATIONAL SEARCH REPORT

International Application No.  
PCT/US 99/20084

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 7 C12N15/12 C07K14/705 C07K16/28 C12Q1/68 G01N33/53 G01N33/68 A61K38/16 A01K67/027		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 7 C07K C12N C12Q G01N A01K A61K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	STRAUSBERG ET AL.: "National Cancer Institute, Cancer Genome Anatomy Project" EMBL DATABASE ACC. NO.: AA883367, 30 March 1998 (1998-03-30), XP002125730	1,3,4
Y	the whole document	2,5-31
Y	STADEL J M ET AL: "Orphan G protein-coupled receptors: a neglected opportunity for pioneer drug discovery" TRENDS IN PHARMACOLOGICAL SCIENCES,GB,ELSEVIER TRENDS JOURNAL, CAMBRIDGE, vol. 18, no. 11, page 430-437 XP004099345 ISSN: 0165-6147 page 434, column 2 -page 436; figures 2-4  -/-	2,5-31

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

**\* Special categories of cited documents:**

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "A" document member of the same patent family

Date of the actual completion of the international search

17 December 1999

Date of mailing of the international search report

11/01/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 661 eport,  
Fax (+31-70) 340-9018

Authorized officer

van Klompenburg, W

# INTERNATIONAL SEARCH REPORT

Int. Patent Application No.  
PCT/US 99/20084

## C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>HILLIER ET AL.: "The WashU-Merck EST Project"  EMBL DATABASE ACC. NO.: R35286,  4 May 1995 (1995-05-04), XP002125731  the whole document</p>	1,3,4
A	<p>EP 0 787 797 A (SYNAPTIC PHARMA CORP)  6 August 1997 (1997-08-06)  page 14, line 20 -page 21, line 5; claims  1-65; table 1</p>	1-31



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 99/ 20084

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
Remark: Although claim 26  
is directed to a method of treatment of the human/animal  
body, the search has been carried out and based on the alleged  
effects of the compound/composition.
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such  
an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this International application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all  
searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment  
of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report  
covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is  
restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Int. Serial Application No

PCT/US 99/20084

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0787797 A	06-08-1997	US 5155218 A	13-10-1992
		AT 171211 T	15-10-1998
		AU 657686 B	23-03-1995
		AU 7879891 A	27-11-1991
		CA 2082390 A	09-11-1991
		DE 69130219 D	22-10-1998
		DE 69130219 T	12-05-1999
		DE 530265 T	31-03-1994
		EP 0530265 A	10-03-1993
		ES 2059298 T	16-11-1994
		FI 925034 A	06-11-1992
		JP 6502297 T	17-03-1994
		NO 924289 A	18-12-1992
		WO 9117174 A	14-11-1991
		US 5786157 A	28-07-1998
		US 5935925 A	10-08-1999